Transit-Oriented Development and Land Use

ROBERT CERVERO
Department of City and Regional Planning,
University of California, Berkeley, CA, USA

Article Outline

Glossary

Bus rapid transit High-quality bus services that mimic some of the mobility benefits of exclusive-guideway rail transit services but at a fraction of the cost, including busways, dedicated-lane services, and buses that operate with special features like signal prioritization and queue-jumper lanes at busy intersections.

Joint development Development activities on or near a transit agency's property that generates revenue and ridership benefits.

Public transportation Common-carrier forms of mass transportation available as a mobility service to the general populous, including heavy rail, light rail, and bus transit.

Self selection Willful decisions of residents to move into particular neighborhoods for lifestyle reasons and thereby have a bearing on travel behavior.

Smart growth Compact, mixed-use urban development that encourages alternatives to automobile travel, including transit usage, and that promotes environmental conservation and which can include, but is not limited to, transit-oriented development.

Transit value capture Securing profits from increases in land values conferred by public transit investments to help finance capital costs, including fixed guideways, rolling stock, and neighborhood improvements.

Definition of the Subject

Transit-oriented development, or TOD, is a term used to describe the physical integration and linkage of public transportation investments and urban land development at or near a station. TOD typically takes the form of compact, mixed land-use development with high-quality pedestrian environments that is concentrated around rail transit nodes [1]. TODs can also occur around the stations and stops of busways and bus services that operate on exclusive, dedicated lanes, often referred to as Bus Rapid Transit (BRT).

The chief objective of TOD is to encourage larger shares of motorized trips to be taken by public transportation and as a result to reduce traffic congestion and improve environmental conditions, including air quality (since the private automobile is the primary source of ozone, carbon monoxide, and particulate matter emissions in many cities of the world). Secondary objectives of TOD might include revitalizing a long-distressed urban district, encouraging the production of more affordable housing and curbing the consumption of agricultural land and open space.

Introduction

TOD is often viewed as an antidote to car-dependent sprawl. By attracting a mix of residences, businesses, shops, and civic activities within a quarter-mile walking distance of an urban railway station, proponents argue that TOD can draw people to transit and thus relieve traffic congestion, improve air quality, and contribute toward climate stabilization. The station and its immediate surroundings can also serve as the hub of a community – a focal point for organizing the regeneration of stagnant neighborhoods and in the case of greenfields, the design and construction of new ones. Thus TOD is championed for both its environmental and its place-making benefits. Of course, the two are not

M. Ehsani et al. (eds.), *Transportation Technologies for Sustainability*, DOI 10.1007/978-1-4614-5844-9,

Originally published in
Robert A. Meyers (ed.) Encyclopedia of Sustainability Science and Technology, © 2012, DOI 10.1007/978-1-4419-0851-3

unrelated. Traffic-snarled districts hardly make attractive, livable places. And as discussed later, places that are conducive to transit riding can help free-up traffic jams.

TOD has gained popularity in the United States, the world's most car-dependent country, for three key reasons. One, it is arguably the most cogent, understandable form of "smart growth." Citizens, politicians, and city-planners alike understand that if there is a logical place to concentrate urban growth, it is in and around transit stations. Second, demographic and lifestyle trends are working in favor of TOD. Living around transit stations appeals to growing numbers of Americans, like childless couples, Generation X'ers, and empty-nesters who value convenience and access to high-quality transit and who place a premium on being in a walkable community with urban amenities. Recent studies have rated TOD as a top real estate investment, estimating that as many as one- third of newly formed households in rail-served US metropolitan areas are receptive to transit-oriented living [2]. For many, TOD residency is a quality-of-life issue. As one study put it, "TODs are all about creating choice for people, about housing, lifestyle and travel"[3]. Third, more and more public policy initiatives favor TODs. As the federal, state, and local government levels, funding programs and incentives in the form of regulatory relief are favoring public and private invests in and around transit stations.

The Scope of TOD

A national study in the United States found that in 2003 about 100 of the nation's 3,300 fixed rail transit stations, or around 3%, could be considered TODs – in the sense all had a mix of land uses, embodied urban designs that created safe and attractive walking environments, and met minimum density thresholds (at least 12 dwelling units and 50 workers per net acre of land) [4]. Far more common than TODs, however, are what have been called "Transit adjacent developments," or TADs. TADs comprise buildings that are near transit stops but fail to induce residents, workers, and shoppers to patronize transit because of their designs and physical orientations. A mid-rise office building with standard parking supplies whose main entrance faces away from the nearby train station, forcing patrons to take a round-about path, is more TAD than TOD. By one definition, TAD is "physically near transit (but) fails to capitalize upon this proximity…(It) lacks any functional connectivity to transit – whether in terms of land-use composition, means of station access, or site design" [5].

The trade-off between creating a TOD versus a TAD reflects the inherent conflict between the transit station as "nodes" versus "places." On the one hand, stations are logistical nodes wherein cars, buses, taxis, delivery trucks, pedestrians, and cyclist converge for accessing transit and allowing intermodal transfers. On the other hand, stations and their environs are places for creating or rebuilding community hubs. Whenever the logistical needs of a station win out, the resulting road designs and parking layouts often detract from the quality of walking, creating more of a TAD than a TOD. This has often been the case in and around suburban rail transit stations in the United States.

On the global stage, TOD is most fully developed in Europe, and in particular Scandinavia. In cities like Copenhagen, Denmark, and Stockholm, Sweden, corridors for channeling overspill growth from the urban centers were identified and reserved early in the planning process, and rail infrastructure was built, often in advance of demand, to steer growth along desired growth axes [6]. As importantly, greenbelt wedges set aside as agricultural preserves, open space, and natural habitats were also designated and accordingly major infrastructure was directed away from these districts. In the case of Stockholm, the last half-century of strategic regional planning has given rise to a regional settlement and commutation pattern that has substantially lowered car-dependency in middle-income suburbs. Stockholm planners focused on creating jobs-housing balance along rail-served axial corridors. This in turn has produced directional-flow balances. During peak hours, 55% of commuters are typically traveling in one direction on trains and 45% are heading in the other direction [6]. Moreover, Stockholm's transit modal share is nearly twice that found in bigger rail-served European cities like Berlin and even higher than inner London's market share. In fact, Stockholm is one of the few places where automobility appears to be receding. Between 1980 and 1990, it was the only city in a sample of 37 global cities that registered a per capita decline in car use – a drop off of 229 annual kilometers of travel per person [7].

Bus-Based TOD: Bogotá and Curitiba

While rail-based TODs are most prevalent worldwide, in good part because rail transit delivers the most regional accessibility benefits and thus are attractive to private investors, in recent time TODs are also taking form along Bus Rapid Transit (BRT) corridors. Many medium-sized global cities are looking to BRT as the most affordable form of high-performance public transit investment. BRT aims to achieve the speed and performance advantages of grade-separated services at a fraction of the cost by cleverly using bus-based approaches. Among its key features are: exclusivity, notably physical segregation; seamless (same-level) transfers; advanced bus technology: clean fuels, lightweight materials, low floors, advanced communications, docking systems; supportive armature: signal priorities, bus turnouts, curb realignments, automated vehicle location (AVL) systems, automated routing and dispatching; and expeditious fare collection and boarding: off-vehicle payment, smart cards. Two noteworthy international experiences with BOT and TOD, both in Latin America, are Curitiba, Brazil and Bogotá, Colombia.

The environmental benefits of balancing urban growth along bus-served linear axes and aggressively pursuing a "transit first" policy are underscored by experiences in Curitiba, Brazil. Curitiba, widely viewed as one of the world's most sustainable, well-managed metropolises, is also one of the most accessible – a product of some 40 years of carefully integrating urbanization and transportation improvements. By emphasizing planning for people rather than cars, Curitiba has evolved along well-defined radial axes that are intensively served by dedicated busways. Along some corridors, streams of double-articulated buses haul 16,000 passengers per hour, comparable to what much pricier Metrorail systems carry. A design element used to enhance accessibility in Curitiba is the "trinary" – three parallel roadways with compatible land uses. An important benefit of mixed land uses and transit service levels along these corridors, besides phenomenally high ridership rates, has been balanced, bi-directional flows, ensuring efficient use of available bus capacity, just as in the case of Stockholm. On a per capita basis, Curitiba is one of Brazil's wealthiest city yet it averages considerably more transit trips than much bigger Rio de Janeiro and São Paulo [6]. It also boasts the cleanest air among any Brazilian city over 1 million inhabitants, despite being a provincial capital with a sizable industrial sector. The strong, workable nexus that exists between Curitiba's bus-based transit system and its mixed-use linear settlement pattern deserves most of the credit.

Bogotá, the Andean capital of Colombia, has gained global recognition for its highly efficient and productive bus rapid transit (BRT) system. For a city of 7.5 million inhabitants facing civil conflict and deep economic problems, Bogotá's emergence as one of the world's most sustainable metropolises is all the more remarkable. In the late 1990s, Bogotá began operating a high-speed, high-capacity bus system, called Transmilenio, building upon Curitiba's successes with dedicated busways. A big difference, however, is that Curitiba relies principally upon circular, crosstown bus routes to interconnect radial busways. Outside of downtown, relatively little was invested in pedestrian and bikeway improvements. Bogotá, on the other hand, actively embraced pedestrian and bicycle access.

The 42-km, 3-line Transmilenio busway is the centerpiece of Bogotá's vast bus network. (The dedicated system will eventually expand to 22 lines covering 391 km.) Bus lanes are situated in boulevard medians, with weather protected, attractively designed stations spaced every 500 m or so. Because of dual carriageways that enable buses to overtake each other and high-level platforms that allow expeditious boardings and alightings, Transmilenio has a throughput of some 35,000 persons per direction per hour, a number that matches that of many Metrorail system. Some one million passengers ride Transmilenio buses each weekday, three times the ridership of two rail lines in Medellin, Colombia (achieved at less than one-fifth of the Medellin Metro's construction costs) and according to one study providing for a social rate of return of 61% [8].

Particularly important to the transitway has been Bogotá's attention to pedestrian and bicycle access, in the form of "green connectors." Perpendicular and grade-separated pedways and bikeways connect some of the poorest barrios and informal housing settlements (with highly transit-dependent populations) to the busways. Other innovative features of Bogotá's sustainable transport program include license-plate rationing, parking management, and car-free

districting. Bogotá is an extraordinary example of matching infrastructure "hardware" with public-policy "software." The city boasts Latin America's most extensive network of cycleways (250 km), the world's longest pedestrian corridor (17 km), and the planet's biggest Car Free Day (covering an entire city of 35,000 ha).

While Bogotá's Transmilenio system has not dramatically altered the cityscape to date, at least when compared to places like Curitiba, research shows that commercial properties have reaped benefits from proximity to busway stations. A hedonic price- model analysis found a monthly rental discount of 1.87% for every additional 0.1 km from a BRT station, all else being equal [9]. This suggests a pent-up market demand for the accessibility benefits conferred by high-quality bus-based transit in cities of developing countries. A more recent analysis found that land-value capitalization increased as Transmilenio expanded into new neighborhood, revealing the network benefits of adding regional connectivity to the system [10].

As with many successful transit investments, it has been the attention to design details, matched by good macro-scale planning, that has contributed to Tranmilenio's success [11]. Car parking is mainly limited to the end stations of the Transmilenio busway. Nearly half of the 57 intermediate stations are served by skywalks/pedestrian overpasses. A phalanx of sidewalks and bikeways feed into all stations, most embellished by vegetative landscaping. Some two dozen civic plazas, pocket parks, and recreational facilities lie within a half kilometer of busway stops. Today, an estimated 45% of Transmilenio users reach stations by foot or bicycle [11].

America's TOD Success Story: Arlington County, Virginia

No place in the United States has witnessed more mid-to-high rise, mixed-use development along a rail corridor over the past three decades than Arlington County, Virginia [4]. This 26 square mile county just south of the nation's capital has experienced a tremendous increase in building activity since Washington Metrorail's 1978 opening: more 25 million square feet of office space, 4+ million square feet of retail space, some 25,000 mixed-income dwelling units, and over 6,500 hotel rooms. Of the nearly 190,000 people today living in Arlington County, in the early 2000s, 26% resided within a Metrorail-served corridor (roughly a one-quarter mile walkshed of stations), even though these corridors comprised only 8% of county land area. If the development added to these two corridors had been built at suburban density standards, such as in neighboring Fairfax County, Virginia, seven times as much land area would have been required [4].

Arlington County's TODs have hardly been the result of good fortune or happenstance. The transformation of once-rural Arlington County into a showcase of compact, mixed-use TOD has been the product of ambitious, laser-focused station-area planning and investment. Prior to Metrorail's arrival, Arlington County planners understood that high-performance transit provided an unprecedented opportunity to shape future growth and proceeded to introduce various strategies – targeted infrastructure improvements, incentive zoning, development proffers, and permissive and as-of-right zoning – to entice private investments around stations. After preparing countywide and station-area plans on desired land-use outcomes, density and setback configurations, and circulation systems, zoning classifications were changed and developments that complied with these classifications could proceed unencumbered. The ability of complying developers to create TODs "as-of-right" was particularly important for it meant developers could line up capital, secure loans, incur upfront costs, and phase-in construction without the fear of local government "changing its mind." Another key factor was the decision not to align Arlington County's Metrorail rail corridor in the median of Interstate-66, thus suppressing development potential. Instead, County officials persuaded the region's transit authority to align the corridor in the traditional urban centers to help jump-start the process of urban regeneration.

The pay-off of concentrated growth along rail corridors is revealed in Arlington County's transit ridership statistics. The County today boasts one of the highest percentages of transit use in the Washington, DC. region, with nearly 40% of Metrorail corridor residents commuting to work by public transit [4]. An important outcome of promoting mixed-use development along Arlington County's rail corridors has been balanced jobs and housing growth which in turn has produced

balanced two-way travel flows. Counts of station entries and exits in Arlington County were nearly equal during peak hours as well as the off-peak. During the morning rush hours, many of the county's Metrorail stations are both trip origins and destinations, meaning trains and buses are full in both directions. The presence of so much retail-entertainment-hotel activities along the County's Metrorail corridors has further filled trains and buses during the midday and on weekends. Arlington County averages higher shares of transit boardings and alightings at its stations in off-peak hours than other jurisdiction in the region with the exception of downtown Washington, DC.

The Transportation and Environmental Benefits of TOD

The litmus test for whether TOD yields environmental and other societal benefits is the degree to which it increases transit ridership, and in particular draws former motorists into trains and buses. Past research in the United States shows that neighborhoods designed according to TOD principles are associated with lower car ownership rates [4, 12], appreciably higher transit modal splits for commuting [13, 14], and fewer vehicle trips per day [15]. In California, surveys show that residents who live near a transit station use transit for their commutes at a rate four to five times higher than residents of the same region who don't live near stations [14]. This pattern has held steady over time. In the case of the Pleasant Hill station of the Bay Area Rapid Transit (BART) system in the San Francisco-Oakland region, for instance, 47% of station-area residents took transit to work in 1993 [13]. Ten years later, in 2003, the share was 44% [14].

While TODs bump up the shares of residents' trips by transit, more important to many observers – particularly traffic engineers and elected officials – is the impact on local traffic conditions. A recent study of 17 TOD projects in five US metropolitan areas (Philadelphia, Northeast New Jersey, Washington, DC, Portland, and the San Francisco Bay Area) uncovered evidence of "vehicle trip de-generation" [15]. Over a typical weekday period, TOD housing projects in these five regions averaged 44% fewer daily vehicle trips than that estimated by the *Trip Generation* manual of the Institute of Transportation Engineers (ITE), which is widely used as the reference manual for gauging traffic impacts of new projects. During peak hours, transit-oriented housing projects generated one-half the typical number of vehicle trips per dwelling unit.

The higher transit use and lower trip generation rates among station-area residents are largely a product of what economists call "self selection." People who prefer to take a train to work – whether to avoid the stress of fighting traffic or to have time to read a newspaper en route to work – purposely choose to live near a rail stop, that is, they are predisposed to transit commuting; they are not "converted" to transit use simply because of where they live. Residing near transit is thus a lifestyle choice. A recent study estimated that around 40% of the ridership bonus attributed to TOD is due to self selection [16].

While TODs generate fewer trips per dwelling unit, this does not necessarily mean they reduce local traffic congestion. Indeed, often the opposite holds which poses a fundamental dilemma – not only for TODs but for all urban developments that increase average densities. Invariably, because not all TOD residents take transit, denser development will congest the nearby road intersections during peak periods. Experiences show that in settings like the United States, where the majority of households own cars, a given percentage increase in density (dwelling units per acre) will not be matched a similar percentage decline in car trips per acre, meaning that car-traffic densities invariably rise. Thus in the near term, even if it is well served by transit, dense development translates into more traffic congestion. Not-in-my-backyard (NIMBY) opposition to higher density has stopped TOD plans around a number of middle-income neighborhoods served by the heavy-rail BART system in the San Francisco East Bay, mainly through building height restrictions. Over the past 20 years, some 2,400 new households have located within a 5-min walk to the Pleasant Hill BART station. Once a critical mass of TOD residents forms, they create neighborhood associations which, among other things, stop efforts to add more development. In Pleasant Hill, plans to build a massive entertainment complex, with a 20-screen IMAX movie theater, were scuttled due to neighborhood backlash. The lesson is clear: do not expect to transform middle-income, stable neighborhoods with large-scale infill projects, because residents have the

political might to stop such proposals every step of the way. More acceptable are TOD proposals for transitional neighborhoods, redevelopment districts, or greenfields with few people around to oppose new development. Also, large-scale, regional trip generators that bring outsiders to neighborhoods made up of professional-class residents will rarely be embraced by residentially dominated TODs. Only local- and neighborhood-serving land uses will be welcomed in such settings.

The downside of preventing new growth around transit stations is that the development will end up elsewhere and most likely increase vehicle miles traveled (VMT) – the strongest single correlate of greenhouse gas emissions (GHG), air pollution, and energy consumption in the transport sector. Like toothpaste in a tube, if development is squeezed from one area, it simply gets redistributed elsewhere. Shifting growth from TOD to AOD (automobile-oriented development) can reduce local traffic congestion at the cost of environmental degradation for a region at large.

It can be argued that there is good and bad congestion, just like good and bad cholesterol. Congestion caused by TOD, many urbanologist would contend, is, for the most part, good. There is no avoiding the fact that denser areas have denser traffic. Dense cities with world-class rail systems, like Paris and London, are often more traffic-choked than sprawling cities. Yet these dense cities are also attractive places to live, work, and visit. Congestion is part of the territory of being an active, vibrant place. TOD, moreover, offers a relief valve to congestion. In a TOD, it is easier to predict when and where congestion will occur, and it is easier to avoid that congestion, e.g., one can hop on a train or ride a bike.

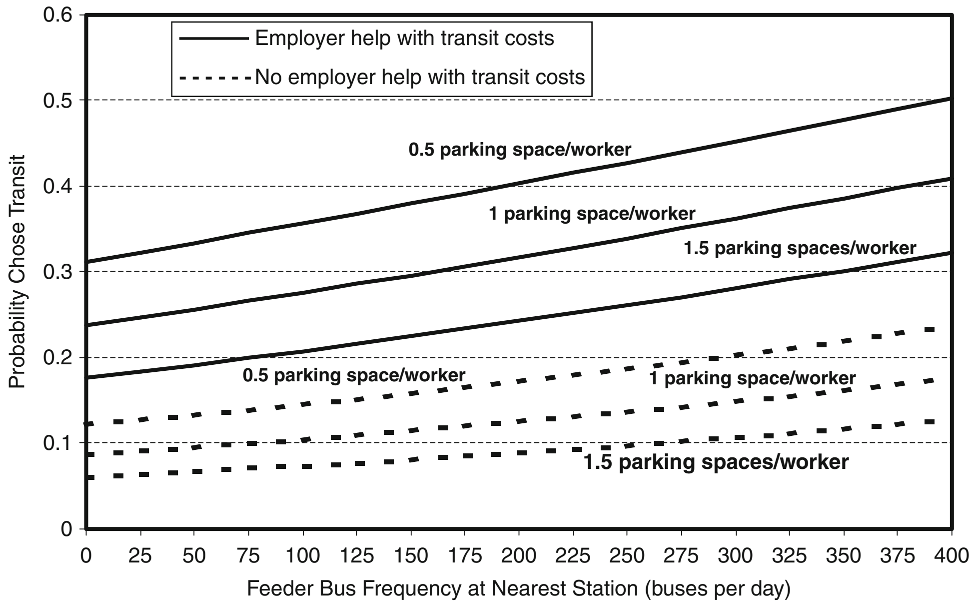
TOD and Ridership at the Workplace

While TOD planning and design has traditionally focused on residential development, unless nonresidential destinations are also served by high-quality transit, few TOD residents opt for rail or bus travel. Studies show that TOD workplaces also yield ridership dividends, though not because of self-selection but rather proximity and convenience [17]. Parking policies and designs are particularly important determinants of whether those working in TODs ride transit or not.

A study of workers in 10 office buildings near suburban rail stations in California found, on average, 19% of commute trips were by public transit, with considerable variation [14, 16]. Other factors, however, weighed heavily on whether TOD workers commuted by transit. This is reflected by the logit-model results in Fig. 1. The figures show the likelihood that TOD workers in California commuted by transit as a function of parking supplies, employer assistance with transit costs, and frequency of feeder bus services. With 25 feeder buses per day, an office setting with 50% more parking spaces than workers, and no employer help with transit costs, the model predicts that just 8% of office workers near a rail station will commute by transit. At the other extreme, for a worker heading to a station with 400 daily feeder buses who work for an employer who provides transit-pass assistance and provides one parking space for every two workers, the likelihood that he or she will commute by transit is 50%. Over the range of feeder bus frequencies, the differential in transit commuting probabilities is 30–40% depending on how generous employers are in promoting transit (i.e., minimal parking and help with transit costs) or in accommodating the automobile (i.e., ample parking and no help with transit costs). Clearly, successful TODs are far more than physical design challenges: the policy "software" that accompanies the built-design "hardware" matters tremendously.

TOD and Parking

Critics charge that many large-scale housing projects near urban rail stations are "over-parked" – more parking is provided than is needed [18]. Excessive parking can drive up the cost of housing, consume valuable land near transit, and impose such environmental costs as increased impervious surface areas (that contribute to urban heat-island effects and stream-water pollution from oily run-offs). As noted, transit stations are thought to offer "location efficiency," enabling residents to get by with fewer cars than they might otherwise own. Yet lenders and planners often insist that code-standard parking be provided in station areas regardless. This can drive up the cost of station-area housing, thus defeating the purpose of promoting affordable housing construction near major transit stops.



Transit-Oriented Development and Land Use. Figure 1
Influences of employer policies and feeder bus frequencies on transit commuting among office workers in California

A recent study raises some questions whether suburban TODs indeed average less parking, at least in the United States [15]. In the case of Portland, Oregon, while average vehicle trip generation rates of TOD projects were 41% below ITE rates, the average use of parking spaces was only 11% less. The parking occupancy at three of the 15 surveyed TOD projects in Portland was actually higher than that predicted by the ITE *Parking Generation* manual. In the San Francisco Bay Area, owning and parking a car was even more of a necessity. There, TOD parking rates were equivalent to ITE's standard of 1.2 spaces per unit. For all seven TOD housing projects surveyed near the Fremont BART station, parking levels were actually higher than the ITE rates – as much as 40% above.

What might explain high parking rates combined with high transit ridership rates for TODs? One factor may be that most cities do not reduce the parking requirements for TODs which in turn induces car ownership. One survey of TODs in California found no reduction in cities' parking requirements at seven of the 11 sites studied [17]. Planners appear to assume that more transit will not reduce parking demand, and they may be right. In most suburban TODs, many residents still need access to a car. They just do not use them as much. But like most suburbanites, they still need a car to get to most non-work destinations – the vast majority of which are located away from rail stops. While transit-oriented housing might mean that more trip *origins* are near rail stops, as long as most *destinations* are not, TOD residents will still own cars and use them for shopping, going out to eat, and the like.

The fact that some suburban TODs encourage transit riding yet fail to prompt residents to reduce car ownership suggests that TODs might be a natural setting for carsharing. A study of City CarShare members in San Francisco found that 4 years into the program, 29% of carshare members had gotten rid of one or more of their cars and 63% lived in zero-vehicle households [18]. Putting shared-cars in and around TODs could relieve many households from owning a second car or a vehicle altogether. Through a combination of proximity advantages and lifestyle predispositions, living near transit can *de*-generate vehicle trips. And with the option of carsharing, it might reduce parking demands as well. This has been the case in cities like Zurich, Switzerland. Zurich, for example, has the second highest per capita transit use in the world (over 600 transit trips per capita per year) and the highest per

capita carsharing participation anywhere (7% of households are members of Mobility Carsharing Switzerland) [6]. And in spite of having one of the world's highest per capita incomes (on a purchasing power parity basis), only one out of three households has an off-street parking space. Zurich also has the highest commercial real estate prices in Europe (along Bahnhoffstrasse) and according to the management consulting firm, Arthur D. Little, ranks number one in quality of life among global cities [6]. In Zurich, world-class transit, carsharing, parking limits, and prosperity go hand-in-hand.

Another policy response might be the introduction of more flexibility in parking policies for housing near rail stops. Flexibility can be in the form of enabling projects to provide below-code parking levels when justified, e.g., compact projects with short, direct walking connections to transit and perhaps on-site retail establishments. Flexible parking standards provide latitude in providing the optimal number of parking spaces [19]. Flexibility can also take the form of unbundling the cost of providing parking from the cost of building (or renting) housing [20]. This would allow developers to better scale the amount of parking provided to what each tenant or homeowner is willing to pay for each car owned, i.e., let the market demand, rather than a possibly out-dated government fiat, determine supply. And flexibility can be in the form of allowing TOD tenants to choose deeply discounted transit passes for frequent riders instead of a 300 square foot parking space.

TOD Implementation Tools

Moving from the theory to the practice of TOD is often fraught with difficulties, especially in car-dependent countries like the United States. A national survey of urban planners in US cities with TODs cast light on the kinds of implementation tools being used to leverage TOD [4]. Most prominent has been zoning, usually in the form of overlays. Overlay zones are frequently introduced on an interim basis to head-off auto-oriented uses that might compromise a TOD and to specify desired land uses as of right, such as housing and convenience shops. For urban TODs, densities of 20–30 dwelling units per residential acre and floor area ratios (i.e., building area divided by land area) of 1.0 and above are not uncommon. Some of the more progressive TOD zoning districts, such as those found in Portland, Oregon, Seattle, San Diego, and Denver, also lower requirements for car parking and sometimes even for bicycles. The city of San Diego, for instance, recommends parking reductions as high as 15% for urban TODs.

Besides zoning, other tools frequently used in the United States to encourage TOD include: funding for station-area planning and ancillary capital improvements; density bonuses, sometimes used to encourage affordable housing; and relaxation of parking standards. Next in the order of frequency of usage have been land-based tools, like land purchases on the open market (for land-banking and potential "deal-making") and assistance with land assemblage. For the most part, redevelopment agencies have applied these tools, meaning their role in leveraging TOD has been mainly limited to economically depressed or blighted neighborhood settings. Because of the higher risk involved, redevelopment tools have often been accompanied by other funding sources, sometimes with a dozen or more participants involved in the process.

Central-City Redevelopment Around Streetcar Lines

A good example of successfully using zoning tools and financial incentives to leverage central-city redevelopment in a transit corridor is the Pearl District in downtown Portland, Oregon. The Pearl District is the most dramatic transformation of downtown Portland in the last 20 years. Once home to a large artist community and an "incubator" for start-up businesses in abandoned warehouses, the Pearl District is now an emerging mixed-use neighborhood of upscale loft housing, parks, art galleries, boutiques, cafes, and restaurants. Since 2001, more than 1,700 condominiums and apartments have been built. Moreover, 55% of all development in downtown Portland has been within a block of the streetcar service since its opening [21].

A major catalyst to the transformation of the Pearl District was the construction of the Portland Streetcar, the first modern streetcar system to be built in the United States. The streetcar has been equal parts housing and transportation tool, as streetcar construction

was explicitly linked to high-density development via density bonuses and an innovative developer agreement. As a result of this agreement, the average density of the District is now 120 housing units per acre, the highest in the city. Moreover, properties near the streetcar line more closely approach permissible densities than properties situated farther away [21].

Tax incentives were introduced to moderate possible gentrification effects of streetcar-induced redevelopment. Many affordable housing projects in Portland receive 10-year property tax abatements if they are within walking distance of a rail line. While the abatements are loosely related to projected price levels and affordability, their primary purpose is to ensure denser development (as specified in the developer agreement) than the market would otherwise support.

Residents in the Pearl District fit the demographic profile found in other Portland area TODs – childless, and either young people seeking smaller lofts, older professionals looking for an urban lifestyle with little upkeep ("downsizing boomers"), or retiring seniors. This variety of homeowner types has contributed to the depth of the market.

The Challenges of Mixed-Use TODs

As dense, mixed-use forms of development, many of the barriers to TOD are generic to all forms of compact growth – not-in-my-backyard (NIMBY) resistance, higher risks and costs, and institutional inertia. Still, some of the barriers to smart growth are more pronounced when it comes to TOD. Mixed-use development is one of them. Mixed land uses, a signature feature of TODs, pose a host of difficulties not only in terms of design but also in lining up funding, investors, and contractors. Planners sometimes impose a design template of ground-floor retail and upper-level housing or offices, i.e., vertical mixing – on any and all development proposals within a TOD. Mixed-use projects are much trickier to design, finance, and sometimes lease than single-use ones [12]. Finding the right formula for mixed land uses can be every bit as difficult as rationalizing parking policies. Vertical mixing is particularly problematic. Quite often, the ground-level retail component of mixed-use TODs suffer the most, in part because they are poorly laid out. Ground-floor retail, for example, is doomed to fail unless it opens onto a street with busy foot traffic and convenient car access. Mixed housing-retail projects also pose unique design challenges. Ground-floor retail needs greater floor-to-floor height (typically 15–18 ft) to be marketable, compared with the 8–10 ft between residential floors. This means the entire ground floor, including multifamily areas, must have higher ceilings, which increases project costs. Ground-floor restaurants pose problems such as where to put the exhaust shafts for kitchens. The exact size and location of restaurant space may not be known until leases are signed. Designers must thus allow exhaust shafts to be put in several potential locations, which can reduce net leasable space. And ground-floor restaurants might be unappealing to upper-level residences seeking quiet and privacy in the evening. Local governments need to be sensitive to these issues and focus more on achieving a desired land-use mix within a transit station area as opposed to individual parcels, i.e., pursue "horizontal" neighborhood-scale mixes versus "vertical" within-building mixes.

Financing TODs

For TODs to take form, there must be money to pay not only for transit infrastructure, like stations and parking, but also the armature – civic squares, streetscape enhancements, pocket parks, expanded trunk-line sewage and piped water capacity – that surrounds a TOD station. Joint development is the form of value capture that has been most widely used to finance America's most successful TODs. Under this approach, a public transit agency partners with a private developer to build a real-estate project on land or air rights owned by the transit agency itself. In return, the transit agency receives either revenue (i.e., revenue sharing) or passes on part or all of the costs of rail-station and ancillary construction (i.e., cost sharing). Among the most common forms of revenue-sharing schemes are land leases, air-rights development, and station interface or connection-fee programs. Cost-sharing schemes include sharing construction expenses, incentive-based programs that produce benefits for private financing (e.g., density bonuses), and joint use of equipment like ventilation systems.

As a value capture tool for transit, joint development has the most appeal in settings where a significant

amount of land is available to a transit agency, preferably purchased on the open market at a fairly low cost (which usually means prior to formal announcement of a new railway project). This has been the case in Hong Kong where the city's transit operator, MTR Corporation, obtains more than half of its revenue from ancillary real-estate property development [22]. To the degree joint development is limited to a geographically restricted area, such as one or two small parcels owned by a transit agency, it fails to capture value from a broader area benefiting from new transit services.

Another common way to leverage and fund TODs, especially in the United States, has been Tax Increment Financing, or TIF. Under this scheme, incremental increases in property tax revenues are channeled back into a district to help revitalize distressed neighborhoods. TIF has been used extensively to encourage TOD in many parts of the United States. In the city of Chicago, half of the 129 TIF districts in 2003 contained a railway station [4]. The state of Pennsylvania applies TIF in Transit Revitalization Investment Districts (TRIDs) to promote economic development and TOD. A criticism of TIF is it diverts tax dollars that would otherwise accrue to a city's general treasury and as thus creates a privileged (i.e., subsidized) zone. Such equity concerns have limited the application of TIFs to much of the developing world.

Some TODs have relied on impact fees as their chief source of funding. These are charges assessed on new development to defray the cost of expanding and extended public services. In 1981, the city of San Francisco, California introduced a Transit Impact Development Fee (TIDF) levied against new downtown office buildings to produce income for operating and maintaining the city's MUNI transit network, comprising light rail, diesel bus, trolley bus, and cable-car services. TIDF revenues produce around $10 million annually. Broward County, Florida, has a Transit Oriented Concurrency system that similarly generates income for transit operations by levying a fixed fee on new development. This program covers around 30% of annual bus-transit operating and capital costs for the county [23]. Because limiting impact fees to new development only can deter investments in all but the healthiest real-estate markets and inflates the cost of housing, it has been applied on a limited basis outside of the United States. In the developing world, exactions are more common wherein developers cover the cost of expanding infrastructure (e.g., water and sewer trunkline capacities) to serve a site.

Future Directions

TODs are a type "smart growth" and sustainable urbanism that have tremendous growth potential due to both supportive public policies and market forces. Global experiences show that the integration of public transport and land use can yield appreciable sustainability benefits that redound to both the private and public sectors. In many settings, developers know they can earn handsome profits building, leasing, and selling TODs. As long as traffic congestion continues to worsen in TODs and demographic trends continue to shift in the direction of smaller, non-traditional households, there will always be a market demand for TODs. However, forward-looking and pro-active public-sector strategic planning will always be a necessity as well if sustainable and cost-effective TODs are to take form. In many automobile-reliant cities of the world, TODs often take the form of a few nodes, or islands, in a sea of auto-oriented development. Unless there are enough TODs, linearly aligned along natural travel corridors to create a critical mass, few transit ridership (and correspondingly auto-reducing) benefits will accrue. That is, there need to be enough origin–destination combinations formed by TODs in non-downtown settings to draw enough motorists into the trains and buses so as to put an appreciable dent in car traffic. Otherwise, stand-alone TODs can very much backfire by creating dense nodes in otherwise auto-oriented settings so as to increase traffic congestion along roads feeding into transit stations.

What is likely on the horizon for particularly progressive-minded cities might be the emergence of "Green TODs." Green TOD is a marriage of TOD and Green Urbanism. This combination is, potentially, a potent elixir. TOD works on the VMT-reduction side of shrinking a city's carbon footprint. Green Urbanism reduces emissions and waste from stationary sources, in the form of green architecture and sustainable community designs. In combination they can deliver energy self-sufficiency and zero-waste living. Renewable energy might come from solar and wind as well bio-fuels

created from human waste and other byproducts. Recycling and reuse of materials, community gardens, and low-impact building materials further reduce the carbon footprint of Green TODs (Table 1).

Synergies would like accrue from combining TOD and Green Urbanism. The higher community densities needed to fill the trains and buses that serve TODs at the same time reduce heating and cooling expenses from the embedded energy savings of shared-wall construction. Smart electric grids generated by photovoltaic panels and wind turbines designed as canopies above stations might also power light-rail cars, plug-in hybrids, and electric buses. Additionally, the compact, mixed-use pattern of development around rail stations can support walking, biking, and the use of limited range private vehicles to reach relatively nearby destinations, thus helping to grow such infant-industries as electric vehicles. One could imagine a future of hydrogen-fueling and electric-battery swap depots in a green community focused on a rail node and for which the destinations of many travelers are close by.

Hammarby Sjöstad, a brownfield redevelopment in the city of Stockholm, is an example, *par excellence*, of marrying of TOD and green urbanism. The combination of railway services, carsharing, and bike-sharing has dramatically reduced vehicle-kilometers traveled of Hammarby's residents and correspondingly greenhouse gas emissions and energy consumption. In 2008, 25% of trips by Hammarby's residents were by bicycle or walking and 52% were by transit, much higher than regional averages [24]. Carbon dioxide emission from private cars was 52% less for households in Hammarby Sjöstad than in planned communities of similar income structure in the suburbs of Stockholm [24]. And the design of an energy self-sufficient and low-waste community has further shrunk the project's environmental footprint. Today, residents of Hammerby Sjöstad produce 50% of the power they need by turning recycled wastewater and domestic waste into heating, cooling, and electricity.

The future of TOD, worldwide, is undeniably bright. As long as urban ills like traffic congestion and air pollution – and global concerns like oil dependency and climate change – persist or worsen, TOD will be embraced as a desirable form of urbanism. Rapid urbanization, combined with equally rapid urban and inter-city rapid rail construction, in countries like China could mean more TODs find particular acceptance in former third-world countries that are rapidly industrializing and modernizing.

Transit-Oriented Development and Land Use. Table 1
Forms of emission reductions form green TODs

TOD	Green urbanism
Mobile sources • *Transit design*: World-class transit (trunk and distribution) Station as hub • *Non-motorized access*: (bike paths, pedways) • *Bikesharing/carsharing* • *Minimal parking*: (reduced land consumption, building massing, and impervious surfaces) • *Compact, mixed uses*	Stationary sources • *Energy self-sufficient*: (renewably powered – solar, wind turbines) • *Zero-waste*: (recycle, reuse, methane digesters, rainwater collection for irrigation, and gray water use) • *Community gardens*: (compost, canopies) • *Buildings*: Green roofs Orientation (optimal temperatures) Materials (recycled, low impact)

Bibliography

1. Cervero R (2008) Transit-oriented development in America: strategies, issues, policy directions. In: Haas T (ed) New urbanism and beyond: designing cities for the future. Rizzoli, New York, pp 124–129
2. Renne J (2009) From transit-adjacent to transit-oriented development. Local Environ 14(1):1–15
3. Urban Land Institute, San Francisco (2008) Bay area TOD marketplace. Urban Land Institute, San Francisco, p 1
4. Cervero R et al (2004) Transit oriented development in America: experiences, challenges, and prospects. TCRP report 102. National Academy Press, Washington
5. Cervero R, Ferrell C, Murpy S (2002) Transit-oriented development and joint development in the United States: a literature review, research results digest number 52. Transportation Research Board, Washington
6. Cervero R (1998) The transit metropolis: a global inquiry. Island Press, Washington
7. Kenworthy J, Laube L (1999) An international sourcebook of automobile dependence in cities: 1960–1990. University of Colorado Press, Boulder
8. Consejo Nacional de Politica Económica y Social (2000) Sistema de Servicio Publico Urbano de Transporte Masivo de Pasajeros de Bogotá. Departamento Nacional de Planeación, Documento No. 3093

9. Targa F, Rodriquez D (2004) Analysis of Bogotá's bus rapid transit system and its impact on land development. Carolina Plann J Winter:26–36
10. Rodriquez D, Mojica C (2009) Capitalization of BRT network expansions effects into prices of non-expansion areas. Transp Research Part A 43(5):561–575
11. Cervero R, Sarmiento O, Jacoby E, Gomez L, Neiman A (2009) Influences of built environments on walking and cycling: lessons from Bogotá. Int J Sustain Transp 3:203–226
12. Dunphy R, Cervero R, Dock F, McAvey M, Porter D, Swenson C (2004) Developing around transit: strategies and solutions that work. Urban Land Institute, Washington
13. Cervero R (1994) Transit-based housing in California: evidence on ridership impacts. Transp Policy 1(3):174–183
14. Lund H, Willson R, Cervero R (2006) A re-evaluation of travel behavior in California TODs. J Arch Plann Res 23(3):247–263
15. Cervero R, Arrington G (2008) Vehicle trip reduction impacts of transit-oriented housing. J Public Transp 11(3):1–17
16. Cervero R (2007) Transit oriented development's ridership bonus: a product of self selection and public policies. Environ Plann A 39:2068–2085
17. Bernick M, Cervero R (1997) Transit villages for the 21st century. McGraw-Hill, New York
18. Daisa J (2004) Traffic, parking, and transit oriented development. In: Dittmar H, Ohland G (eds) The new transit town: best practices in transit-oriented development. Island Press, Washington, pp 114–129
19. Cervero R, Golub A, Nee B (2007) City carShare: longer-term travel-demand and car ownership impacts. Transp Res Rec 199:70–80
20. Dunphy R, Myerson D, Pawlukiewicz M (2004) Chapter 7: Ten principles for developing around transit. In: Dunphy R et al (eds) Developing around transit: strategies and solutions that work. Urban Land Institute, Washington
21. Shoup D (2005) The high cost of free parking. Planner's Press, Chicago
22. Cervero R, Murakami J (2009) Rail + property development in Hong Kong: experiences and extensions. Urban Stud 46(10):2019–2043
23. Fogarty N, Eaton N, Belzer D, Ohland G (2008) Capturing the value of transit. Center for Transit Oriented Development, Oakland
24. Brick K (2008) Summary report: follow-up of environment impact in Hammarby Sjöstad. Grontmij AB, Stockholm

True Color Night Vision Video Systems in Intelligent Vehicles

Dirk Hertel
Quincy, MA, USA

Article Outline

Glossary

Automotive night vision A technical system to enhance human night vision in situations where visibility is limited or challenging to the human visual system, for example in situations where the range of distances illuminated by headlights is shorter than the safe braking distance or in situations where the human eye may be too slow to adapt to sudden changes in brightness. It requires an effective human–machine interface, for example, a visual display to make available to the driver real-time additional visual information.

CMOS image sensor An active-pixel image sensor (APS) consisting of an integrated circuit with an array of light-sensitive cells (pixels) and manufactured using complementary metal–oxide–semiconductor (CMOS) technology. Each pixel combines a photodetector with an active amplifier that may perform additional functions such as controlling the pixel response to light and extending the pixel's dynamic range.

Driver assistance system (DAS) A technical system to assist the driver in the task of driving, especially in situations where the human sensory system (visual, auditory) or human attention might fail to recognize a dangerous situation, or when the human response time in the perceptual-motor task of recognizing a dangerous situation then using vehicle controls to take corrective action may be too slow to avoid an accident.

Dynamic range Dynamic range is the ratio between the largest and smallest possible signal values of a changeable quantity such as light. For image sensors and cameras, dynamic range refers to input radiance or luminance signals and is defined as the ratio between the maximum and minimum signals where a camera can simultaneously capture detail carried by small signal variations.

Electromagnetic spectrum, visible (VIS) and near-infrared (NIR) The human eye can detect the visible (VIS) portion of the electromagnetic spectrum ranging from wavelengths of about 380 to 730 nm. Infrared is electromagnetic radiation with longer wavelengths than visual, with NIR denoting the shortest wavelengths of the IR-spectrum between 700 and 1,400 nm.

Wide dynamic range (WDR) Also referred to as high dynamic range (HDR), it covers a family of techniques to extend the dynamic range of image sensors beyond that of an image sensor with a linear relationship between input radiance and digital output signal.

Definition of the Subject and Its Importance

Vision-based driver assistance aims to increase traffic safety at night by enhancing human night vision in situations where visibility is limited or challenging to the human visual system. The night vision cameras described sense the near-infrared radiation of the car's own headlights to extend the range of forward vision, while, at the same time, discriminating traffic-relevant colors of signal lights and signage. Adaptive wide-dynamic-range image sensor technology ensures that scene detail in both deep shadow and bright highlight is captured even in situations where the human eye may be

M. Ehsani et al. (eds.), *Transportation Technologies for Sustainability*, DOI 10.1007/978-1-4614-5844-9,

Originally published in
Robert A. Meyers (ed.) Encyclopedia of Sustainability Science and Technology, © 2012, DOI 10.1007/978-1-4419-0851-3

blinded or too slow to adapt to sudden changes in brightness. The information from the near-infrared and visible spectrum is combined into natural-looking color images and communicated to the driver via on-board visual displays to enable recognition of the traffic situation.

Introduction

Since the late 1990s, there has been research into driver assistance systems employing sensors to perceive the automotive environment and utilize their outputs to create warning and vehicle control signals. Vision-based systems are of particular importance because more than 90% of driving decisions and actions are based on vision. Although the human visual system (HVS) is able to adapt to a vast range of lighting conditions, the rate of visual adaptation is often outpaced by the rate of change in lighting conditions while driving, for example when passing blinding headlights at night. Camera-based vision systems have the capability to aid the driver in critical situations by more rapidly adapting to changes in brightness and intra-scene dynamic range. They may also extend spectral sensitivity into the NIR to utilize the invisible portion of the headlight spectrum or can additionally employ NIR-only headlights that can be used in high-beam mode without blinding other drivers. Since the visual channel has the highest capacity of information transfer compared to other senses, image displays are used as human–machine interface to communicate information to the driver. However, the displayed image must be of sufficient quality to meet minimum requirements of usefulness so that scene objects can be seen, and of naturalness so that scene objects can be instantly recognized. For example, displayed objects with too little contrast will be overlooked while objects with unnatural grayscale and color appearance or distorted perspective may not be recognized in a timely fashion.

Although the majority of research on driver assistance systems focuses on image processing to interpret the information delivered by image sensors, comparatively little attention is given to the task of capturing scene detail. Since the camera forms the first link in the image information processing chain, its performance is critical; if the intra-scene dynamic range, the ratio of the brightest to darkest information-carrying intensity, exceeds the dynamic range of the image sensor, then scene detail will be lost and cannot be recovered by image processing.

A model-based methodology is one that:

- Characterizes the automotive imaging environment (photospace) as wide dynamic range (WDR)
- Defines incremental signal-to-noise ratio (iSNR) as the criterion for reliably capturing WDR scene detail
- Uses the inherent sensitivity of silicon-based image sensors to utilize scene information carried by NIR radiation
- Separates color and NIR information in the image sensor and WDR image signal processing
- Defines the principles for combining luminance information from the visual and NIR spectral ranges with traffic-relevant color information into a natural and useful image for visual display

The Automotive Imaging Application

Reduced visibility at night substantially increases the risk of traffic accidents caused by misjudgment of dark roadways or failure to recognize the presence of pedestrians and obstacles. If an accident does occur, passive safety systems such as seat belts and airbags can reduce the percentage of fatalities to the vehicle's occupant, but only to a limited degree [1]. Any further reduction of traffic fatalities requires the use of active driver assistance systems.

Driver Assistance

Driver assistance can be classified into comfort systems and safety systems. These systems employ sensors that take measurements of the automotive environment, interpret the signals delivered by the sensors, generate warning signals, and, if the driver takes no action, even take over vehicle controls. The distinction between comfort and safety is fluid; parking guidance makes parallel parking more comfortable but also enhances safety by detecting persons and objects not visible to the driver. As an example of active safety, lane-keeping systems analyze sensor signals to determine if the vehicle is leaving the lane or roadway, then generate warning signals and in case of driver inaction steering signals. Other examples are lane-change assistance with blind-spot monitoring, collision avoidance, and driver-drowsiness monitoring.

Vision-Based Driver Assistance Vision-based driver assistance has a high potential of increasing traffic safety. First, more than 90% of the information that is critical for making driving decisions comes through the visual channel. Second, conditions that impair the flow of visual information have a significant impact on traffic safety. Nighttime driving is twice as risky as daytime driving, even though the volume of traffic is often less at night. Nighttime accidents tend to be more severe, with a disproportionate increase of fatalities compared to the volume of traffic. Nighttime accidents more often involve pedestrians or cyclists. The illumination of the roadway may be inadequate. For example, dipped headlights stretch over a fixed range of about 180 ft, but the safe braking distance increases with speed exceeding the visibility range at speeds greater than 30 mph [2].

Video-based driver assistance systems provide additional image information in situations where visibility is poor (night vision) or obstructed (side or rear vision). Applications range from driver assistance with forward and rear view images displayed on a dashboard screen to active safety systems where the video footage is analyzed by software that can issue warnings or trigger vehicle controls. On the front-end of these systems are video cameras that must capture scene details reliably under all circumstances.

Requirements of Vision Systems for Driver Assistance The basic requirements for a vision-based driver assistance system are image quality and reliability.

Image Quality Image quality can be defined as “the degree to which the image can be successfully exploited by the observer” [3]. An image of “high” quality should satisfy three criteria: genuineness, usefulness, and naturalness (GUN-model).

- Genuineness is described as the degree of apparent similarity of the reproduced image and its viewing environment with the external reference, that is, the original image and environment. Genuineness is related to fidelity, for example between a painting and its photographic reproduction.
- Usefulness is how relevant the reproduced image and environment are to the observer and the task activity. Usefulness is related to the discriminability of objects in the image, for example obstacles on a dark roadway.
- Naturalness is described as the degree of apparent similarity between the reproduced image and environment and the viewer’s internal references, that is, memory prototypes. It is related to recognizability of objects in the image, for example whether the obstacles on a dark roadway are pedestrians or deer.

Night vision systems for driver assistance do not require genuineness, but they do require both usefulness and naturalness. Color and near-infrared (NIR) can supply important additional information that increases the usefulness of a night vision image, though NIR sensitivity may affect color saturation and thus reduce the naturalness of displayed images (section “Adding Color Differentiation to VIS + NIR-Sensitive Imager”).

Image information that is interpreted by software has to fulfill the criterion of usefulness. Image information displayed for viewing has to be natural so that it can be interpreted instantly. An unnatural image requires effort by the interpreter and thus can distract, for example, false colors, inverted grayscale (thermal images), or distorted perspective (extreme fisheye).

Reliability The main difference between photography and automotive imaging is that the former is in most cases controlled by a conscious subject – the photographer – who selects the frame, knows what is in the scene, and can avoid extremes (strong light sources). However, when a camera is fixed in a vehicle, frame selection and content are arbitrarily determined by vehicle position, so intra-scene dynamic range may include, for example, approaching headlights, glare from other vehicles, tunnel entrances or exits, or solar glare. Image detail provides information about the objects in the real world, such as their location, shape, size, brightness, and color. If the dynamic range of a camera is too narrow to accommodate the intra-scene dynamic range, important object details will be missing from the corresponding image. Once lost by capture failure, these details can never be recreated with image post-processing [4, 5].

Conclusion: Model-Based Approach The design of automotive vision systems is governed by performance versus cost trade-offs. Therefore, it is important that the design is based on a clear understanding of

the nonnegotiable performance thresholds such as sensitivity, signal-to-noise ratio, and dynamic range. A model-based design can quantitatively define and meet the performance requirements of automotive night vision. This includes models of human color night vision, the automotive night vision environment (photospace), and detection of scene detail (signal-to-noise ratio).

Human Perception of the Automotive Imaging Environment

The human visual system (HVS) is able to adapt to a vast range of natural and artificial lighting conditions. The mechanisms of adaptation are complex and include the combined responses of iris, lens, retinal photoreceptors, plus processing in the retina and visual cortex.

Day and Night Vision

Viewing the environment from a fast moving vehicle means that the driver's visual system must adapt to a visual environment that changes more rapidly than for a pedestrian or stationary observer. Under daylight conditions, such changes requiring instant adaptation include driving into the deep shadow of a bridge or tunnel, emerging from deep shadow into bright daylight, and solar glare. At night, the most challenging situation occurs immediately after passing the blindingly bright headlights of oncoming vehicles and having to adapt to the comparatively dark roadway ahead. These examples illustrate that adaptation of the HVS to instantly changing stimuli is the most critical issue in the automotive environment.

However, the models describing human day and night vision refer to steady states of global adaptation to defined ranges of light levels and characterize the responses of the two types of retinal sensors, rods and cones, to these light levels. Three distinct adaptation states of the human eye have been identified: photopic, mesopic, and scotopic.

- Photopic vision in bright light is dominated by cone receptors of relatively low sensitivity that provide full color at high acuity, which is highest at a central foveal field of about 2°. The spectral sensitivity function for photopic vision is measured as the combined response of the long-, medium-, and short-wavelength-sensitive cones and standardized as the photopic luminous efficiency function $V(\lambda)$ with peak sensitivity at 555 nm (yellowish green).
- Scotopic vision, at very low light levels, relies on highly sensitive rods that provide only a monochrome image with poor acuity at the center of the field of vision but increased acuity toward the periphery. The peak of the scotopic luminous efficiency function $V'(\lambda)$ is at 510 nm, shifted toward bluish green, the so-called Purkinje-shift.
- Mesopic vision is a transition phase between photopic and scotopic vision that is invoked at luminance levels below 10 cd/m^2 and above 0.01 cd/m^2. The mesopic luminous efficiency function has not yet been standardized. Proposals range from a simple weighted average of the photopic and scotopic sensitivity functions to more complex chromatic models that weigh the contributions from the individual cones and yield sensitivity functions with two or three peaks [6]. Mesopic vision thus represents a balance between photopic and scotopic vision where the response of monochrome and color receptors will depend not only on overall light level but also on the scene-dependent local distribution of brightness on the retina. For example, the response within those areas of the retina that are stimulated by bright colored traffic lights is still cone-dominated while the response of retinal areas that are stimulated by dark background objects is rod-dominated.

Due to the presence of artificial lights in the automotive environment at night, the eye's adaptation never quite reaches the scotopic state but remains between or in the photopic and mesopic states. Color vision in very dark scenes is restricted to bright colored lights (traffic, tail, and brake lights), retroreflective signage, and surfaces brightly illuminated by headlights and streetlights. As these are relatively small and isolated by a dark background, they are referred to as "unrelated colors" [7]. Thus, relevant traffic lights are perceived as colored, but most dark objects (foliage, pedestrian's clothes) are perceived as monochrome.

Although the simultaneous dynamic range of the human eye, also known as the instantaneous sensitivity window, is limited to about 10,000:1, adaptation

enables the HVS to adapt over time to a much wider range of luminance levels (Fig. 5). The rate of adaptation of the HVS to changing light levels is critical to traffic safety. Adaptation works with two time constants: fast (transient) adaptation occurs within seconds whereas slow adaptation happens on a much longer timescale, ranging from several minutes up to three quarters of an hour. Fast adaptation is the instant reaction to a sudden change in brightness, for example tunnel entry or exit, and vision is impaired until adaptation is achieved, usually after 1–2 s. Therefore, fast adaptation is of utmost importance in the performance of the HVS during driving. The rate of slow adaptation depends on the direction (dark takes longer than bright adaptation) and the light levels before and after adaptation. For example, the brighter the initial light level and the darker the final light level, the more time is required for slow dark adaptation. Headlight glare from oncoming vehicles constantly interrupts the process of dark adaptation.

Challenges to the Human Visual System

Driving under night conditions is very stressful to the HVS. First, the range of luminance levels is extremely large, ranging from blinding oncoming headlights to dark objects beyond the cone of one's own headlights. The constant change in peak brightness and position of the highlights puts the eye under the constant strain of global and local adaptation so that the retina is constantly registering either photopic or mesopic brightness levels. Color vision is restricted to bright lights and brightly illuminated objects whose position in the field of vision changes rapidly and constantly. The rate of change in stimuli will almost always exceed the rate of transient (fast) adaptation. Visual night driving performance can be evaluated by the responses to three visual subtasks that are defined by questions and evaluated by corresponding parameters: "Can it be seen?" (contrast threshold), "How quickly?" (reaction time), and "What is it?" (recognition threshold) [6]. The ability of the HVS to fulfill these tasks decreases with age (Table 1). The most serious aspect of age-related visual deterioration is the slowing of transient adaptation, making it more likely that the driver is "blind" immediately after sudden changes of brightness. At night, headlight glare can lead to the phenomenon of "black holes," a temporary inability to perceive dark background detail after blinding. Although all drivers experience black holes, older drivers take longer to recover vision after the experience. Many societies have an increasing proportion of older drivers, so the implications of this issue are ever greater.

True Color Night Vision Video Systems in Intelligent Vehicles. Table 1 Visual impact of aging on human vision performance

Parameter	Effect of aging	Visual impact
Sensitivity	Decreases	Less light reaches retina
Color perception	Deteriorates	Loss of blue sensitivity
Rate of dark adaptation	Slows	"Black holes"
Glare	Increases	Loss of dynamic range
Contrast sensitivity	Decreases	Loss of detail
Peripheral vision	Decreases	Restricted

The simulation in Fig. 1 illustrates the loss of sensitivity and shift in color vision due to yellowing with age.

Conclusion: Producing Natural Images Under Mesopic Conditions

Night vision images displayed for the driver have to be rendered "naturally" so that traffic-relevant objects can be quickly seen and recognized. This requires that illuminated dark details such as pedestrians are rendered as a sufficiently bright monochrome image while "unrelated color" objects such as traffic lights are rendered with their correct hue and sufficient saturation (Fig. 2). In other words, the best compromise is to reconstruct a good luminance image then add those unrelated colors that are relevant to the traffic scene. This way, the automotive vision system will mimic the mesopic state of the HVS, but with two key improvements. First, as adaptation to sudden changes in light levels is almost instantaneous, the driver is provided with visual information by the safety system in

a b c

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 1
Simulated change in night color vision with age (**a**) 20, (**b**) 60, and (**c**) 70 [42]

a b

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 2
Same scene showing (**a**) related color at daytime and (**b**) unrelated color at nighttime [8]

situations where the HVS adapts slowly. Second, the displayed luminance image is brighter and more detailed at a wider field of view in comparison to the corresponding rod-based HVS perception of the dark background. Although color cameras do not experience a shift in spectral sensitivity akin to the Purkinje shift, image processing algorithms (such as white balancing) still have to be applied to correct color shifts

in illuminated objects caused by altered illuminant color temperature, for example moving between areas lit by mercury and sodium vapor lamps.

The Automotive Imaging Environment

A possible methodology for estimating the dynamic range of the automotive vision environment is based on photometric and spectral measurements together with the statistical photospace model. In addition, knowledge of the spectral characteristics of colored light sources is important when designing VIS–NIR color night vision systems.

Scene Photometry

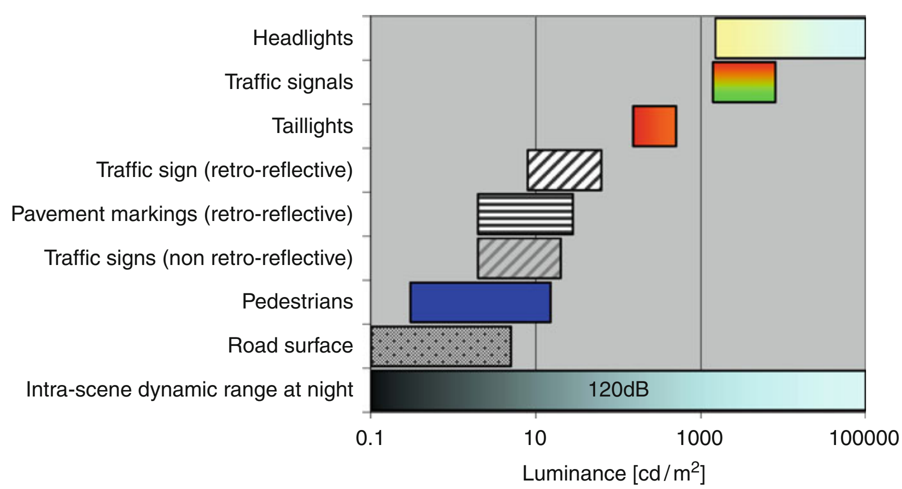
To determine dynamic range requirements, a quantitative model of the image capture environment is needed. First, a measure of the brightness of light sources and illuminated objects in automotive scenes is required. The intensity of light is measured in radiometric power units, which are then converted to photometric units that represent the perceived brightness. This is done by weighting the radiant power at each wavelength by a spectral sensitivity function that models human brightness sensitivity, for example the photopic luminous efficiency function $V(\lambda)$ for the bright-adapted eye, or the scotopic luminosity function $V'(\lambda)$ for the dark-adapted eye. The amount of visible light reflected by or emitted from a surface area is measured as luminance with units of candela per square meter [cd/m^2]. Luminance is a good indicator of how bright a light source or illuminated object will appear. It measures the luminous power perceived at the surface of a light source or an illuminated object. It is independent of viewing distance and not changed by optical systems, e.g., lenses. Luminance measurements of objects in automotive scenes require spot luminance meters.

In the case of color and multispectral imaging, radiance must be measured [W/cm^2/sr] at specific wavelengths to characterize the night vision environment. This is done by a spot spectroradiometer that can measure not only the visible but also the NIR range.

Scene analysis relies on measuring the luminance of individual scene elements rather than the scene average [9]. Scene photometry in combination with scene analysis has been used to build statistical distributions of light levels within a scene, ranging from the brightest lights to the darkest illuminated surfaces. Figure 3 shows an automotive vision example with luminance ranges for light sources and reflective surfaces typical of a night traffic scene.

Photospace in Automotive Vision

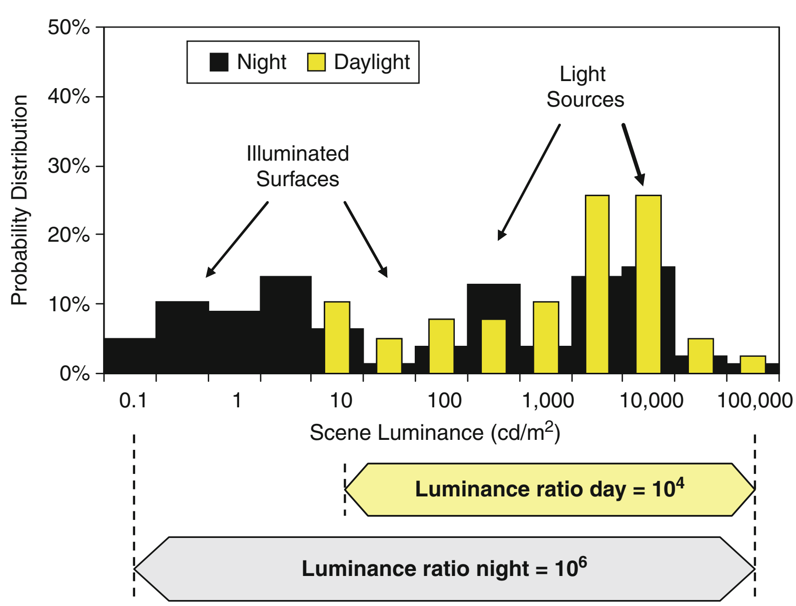
Photospace describes the environment as a statistical distribution that is populated by measuring average



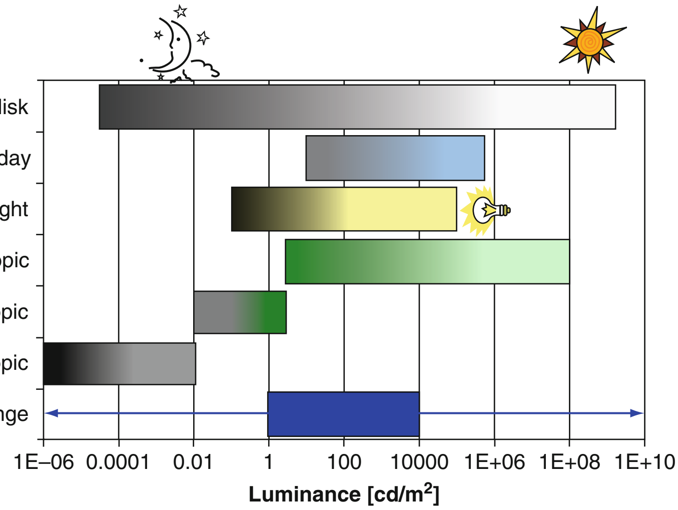
True Color Night Vision Video Systems in Intelligent Vehicles. Figure 3
Luminance ranges for light sources and reflective surfaces in typical night traffic scenes [8]

scene luminance levels for a great number of application-typical scenes [10]. By analyzing at least ten daytime and ten night traffic scenes, extended photospace distributions can be built based not on scene averages but rather on the probability that a scene element of a given luminance will occur (Fig. 4) [4, 11]. As more image data are added, such distributions will ever more closely represent application conditions. These extended photospace distributions for automotive vision by day and night show clusters representing light sources at high luminance levels (reflections of sun, vehicle lights, street lights, and traffic lights), and illuminated surfaces at lower luminance levels (road surface, traffic signs, and traffic participants). The luminance ratio within night scenes (10^6) is anticipated to be higher than that of daylight scenes (10^4–10^5) since poorly illuminated objects at night have much lower luminance. The highest luminance ratios in daylight scenes occur when the roadway enters or exits deep shade (bridges, tunnels). For the camera to detect image detail at night simultaneously in both the brightest (differentiation of headlights) and darkest scene areas (pedestrian detection), a dynamic range of 10^6:1 is required. However, a wide dynamic range is not the only requirement for an automotive camera to work in a night vision environment. The other necessity is that the camera is sensitive to the lowest luminance levels of the automotive photospace. The noise-limited sensitivity of the camera at the video frame rate should be such that its dynamic range covers the entire automotive photospace, including most of the mesopic range plus as much of the photopic range as possible (Fig. 5).

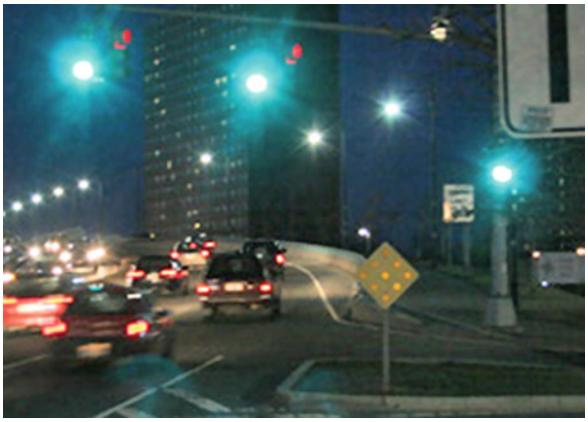
For a monochrome system, it is sufficient to render the brightest lights near the maximum signal level of the pixel in order to minimize halos and flare, thus maximizing differentiation between various lights, for example, headlights. However, in the case of color capture, this may be insufficient; a colored light captured near the maximum signal level will be too desaturated to be recognizable. As a rule of thumb, if a pixel's set of RGB signals is to be perceived as "green" (10% color saturation between the hue angles of 85° and 155°), each of the R and B signals has to be at least 10% lower on average than the G signal. However, if the green light increases in brightness, the G signal will reach then exceed



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 4
Extended photospace distribution showing clusters for light sources and illuminated surfaces [5]



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 5
Luminance levels and adaptation of the HVS [42]



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 6
Typical urban night traffic color scene where limited dynamic range results in clipped highlight colors appearing *white* (centers of *green* traffic lights, *red* tail lights) [35]

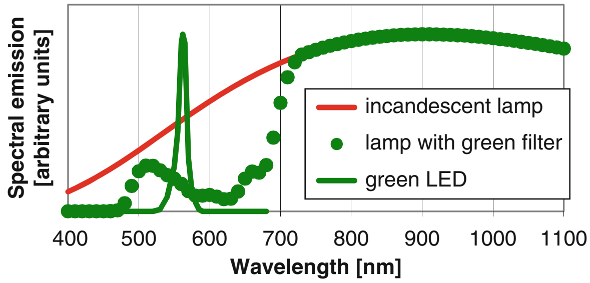
saturation level before R and B, thus decreasing the G/R and G/B ratios so that the green color is barely distinguishable from "white" (see Fig. 6). Therefore, the dynamic range of a night vision color system must be wide enough to prevent any of the RGB color signals from being clipped.

Spectral Distribution of Light Sources

The diversity of spectral characteristics of colored light sources poses a challenge to automotive vision systems that combine NIR sensitivity with color differentiation. For example, a green traffic light might be an LED array or an incandescent lamp behind a colored filter. The former has a narrow spectral distribution peaking in the 500–600 nm range. The latter has a more complicated spectral distribution. Although in the VIS range, it has a peak in the green region of the visible spectrum, since most commonly used color filter materials are almost transparent in the NIR range and the spectral emission of incandescent lamps (color temperature 3,200 K) peaks in the NIR at about 800 nm, the strong NIR contribution might exceed that from the VIS range (400–700 nm), see Fig. 7. Spectral measurements of objects in automotive scenes require spot spectroradiometers.

This high NIR content of colored lights does not pose a problem for human vision and common RGB color cameras because the eye is insensitive to IR, and the latter has an IR-cut filter.

It follows that a VIS–NIR color night vision system has to combine high sensitivity in both spectral areas with good reproduction of those colors relevant to night traffic scenes. This requires the imager to have WDR to avoid signal saturation in colored lights, plus



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 7
Spectral VIS–NIR emission curves of an incandescent lamp (3,200 K), an incandescent lamp with a green glass filter, and a green LED [35]

sufficient separation of NIR and color signals so that colors can be reproduced independently of the light source's IR content.

Conclusion: Automotive WDR Environment

Intra-scene dynamic range is determined by the range of intensities between bright lights (some of them carrying color information) and dark, dimly illuminated objects. The automotive vision environment can be measured by scene photometry to quantify the luminance ranges of traffic-relevant light sources and illuminated objects, and to populate statistical photospace distributions that show the probability that scene elements of a given brightness (luminance) will occur at night or day. Automotive scenes typically cover a WDR, highest at night due to lower background luminance levels. Challenges in daylight conditions are deep shadows (parked vehicle under bridge, tunnel entry) and blinding (solar glare, exit from tunnels or bridges). Night vision requires both high sensitivity (to distinguish shadow detail) and WDR (to distinguish highlight color detail). Preserving the color information of traffic-relevant lights requires an even wider dynamic range than for monochrome vision.

Extending the Dynamic Range

A possible methodology to extend the dynamic range of image sensors without compromising the transfer of scene information to image information can be developed from an information-theoretical model.

Definition of Dynamic Range

Dynamic range generally relates to an imager's ability to capture information about both bright and dark objects in the same scene at the same exposure settings. There are several definitions of dynamic range. In its most general sense, dynamic range describes the ratio between the largest and smallest possible signal values of a changeable quantity such as light [4]. Dynamic range can be either input-referred (ratio of real-world light levels) or output-referred (ratio of the response signals delivered by a light detector). Since the output-referred definition can lead to dynamic range numbers built solely on technical specifications (such as the bit depth of an A/D converter) rather than on actual measurements, it has limited value in predicting dynamic range [13]. Therefore, dynamic range should refer strictly to input signals, for example luminance L in photopic units of cd/m^2. The relationship between input and output signals is established from the imager response curve.

In silver halide photography, the luminance ratio of the highlights to the shadows is called "object ratio" and the logarithm of this ratio "object range" [4]. The equivalent of "object range" in digital imaging is the input-referred dynamic range, which can be expressed in units of decibels [dB] by calculating 10 log of the luminance ratio. However, technical specifications of digital and video image sensors traditionally referred to ratios of output voltages or digitized light levels so that decibels represent 20 log of the ratio (even when the represented input luminance is directly proportional to the output voltage or level) not to its square [14]. In order to avoid disadvantaging input-referred data, the 20 log rule is also used to calculate the input-referred dynamic range from luminance ratios:

$$DR = 20 \cdot \log_{10}\left(\frac{L_{\max}}{L_{\min}}\right). \tag{1}$$

Reliability Criterion

Ascertaining dynamic range requires determining the maximum and minimum signal levels. Detection of the smallest possible signal is limited by the noise floor. To become detectable, a signal must exceed the noise floor; in other words, the signal-to-noise ratio (SNR) must be greater than 1. The largest possible detected signal is

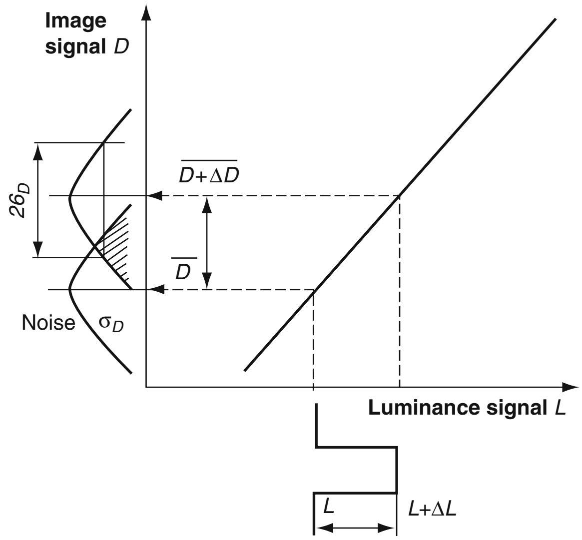
limited by saturation. Any input signal above the saturation level can only reproduce a saturation level output signal and thus remains indistinguishable from the saturation signal. This introduces the concept of signal variation as the carrier of information: object detail is carried by small variations in light level. Scene information can only be gathered reliably if the image capture system can detect and reproduce the small signal variations carrying that information. Thus, dynamic range needs a more precise specification to ensure detection reliability: it is the ratio between the maximum and minimum luminance signals where a camera can simultaneously capture detail carried by small signal variations.

A distinction has to be made concerning the spatial dimension of object details. Below a certain size, the details in an image are no longer formed independently; the signal at a given image point depends not only on the luminance at the corresponding object point but also on that of its surroundings [4]. This effect can be accounted for by the modulation transfer function (MTF). However, this estimation of dynamic range refers only to the reproduction of details large enough that the loss of detail due to MTF can be ignored. It is desirable to deduce dynamic range from objective measurements. This methodology is laid down in a progression of ISO/TC42 imaging performance standards [15–17].

Figure 8 shows the transfer of large-area detail (small luminance variation) by the response curve in the presence of noise [18]. The response curve is the relationship *D(L)* between input luminance signal *L* and output digital image signal *D*. ISO/TC42 refers to the response curve as opto-electronic conversion function (OECF). The slope of the OECF, the change *dD* in image signal due to a change *dL* in the luminance signal, is referred to as incremental gain *g(L)*:

$$g(L) = \frac{dD}{dL}. \tag{2}$$

Detection error is given by noise. Since dynamic range refers to the input luminance signal *L*, the noise σ_L has to be defined as the luminance-equivalent of the noise σ_D that can be observed and measured in the digital image [19]. At each luminance level *L*, the equivalent luminance variation σ_L can be calculated from the image noise σ_D via the incremental gain:



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 8
Transfer of large-area detail (small luminance variation ΔL) by the response curve *D(L)* in the presence of noise σ_D [5]

$$\sigma_L(L) = \frac{\sigma_D(L)}{g(L)}. \tag{3}$$

The criterion for detection reliability is the SNR, the ratio of the input luminance signal *L* and the luminance-equivalent noise σ_L. Detection reliability requires the SNR to stay above a threshold of one to fulfill the following condition:

$$\frac{L}{\sigma_L(L)} = \frac{g(L) \cdot L}{\sigma_D(L)} > 1. \tag{4}$$

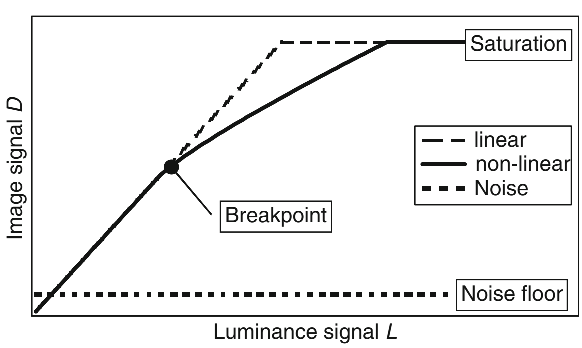
The expression in (4) is referred to as "incremental SNR" (iSNR, also S/Nx and SNR_{inc}); the incremental signal is defined as $iS(L) = g(L) \cdot L$. An equivalent concept for determining detection reliability in silver halide photography is known as noise-equivalent contrast (NEC) [20].

The threshold condition $iSNR > 1$ means that at any given luminance level *L*, detail carried by a small variation in luminance can only be reliably detected if the corresponding incremental image signal is high enough and the noise low enough. Fulfilling the iSNR

criterion will be first explored for imagers with a linear response. At very low luminance levels, noise keeps iSNR below its threshold. The minimum reliable luminance level $L_{\min}$ is reached only when iSNR rises above the threshold. At very high luminance levels, the onset of saturation will clip further signal increases and thus cause the incremental gain to drop sharply. The maximum reliable luminance level $L_{\max}$ is reached when iSNR once more drops below the threshold. It follows that dynamic range for linear imagers is bounded by the two extreme luminance levels $L_{\min}$ and $L_{\max}$ at which the iSNR condition of (4) still holds.

The iSNR condition is now applied to imagers with extended dynamic range. This requires shifting saturation to higher input signal levels by a nonlinear response curve: at a so-called breakpoint, the response curve's slope (incremental gain) is suddenly lowered to delay saturation to higher luminance levels than for a linear imager (Fig. 9). However, lowering incremental gain will, in turn, decrease iSNR, pointing to a potential iSNR trade-off when increasing dynamic range. Rather than just using iSNR to determine the boundary luminance levels $L_{\min}$ and $L_{\max}$, the minimum iSNR criterion has to be maintained at every luminance level in between. Verifying this requires iSNR be measured as a function of luminance L. An equivalent strategy was used for silver halide photography where NEC curves were measured as function of exposure [20].

Therefore, nonlinear imagers require a more general definition beyond the recommendations of ISO/TC42; dynamic range is the input luminance range within which the threshold criterion for



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 9

Extending dynamic range with a nonlinear response curve [5]

$iSNR(L)$ is met at all levels $L_{\min} \leq L \leq L_{\max}$. This suggests a strategy for maximizing the dynamic range of imagers with a nonlinear response curve: maintain the necessary minimum of $iSNR(L)$ over the largest possible range of luminances L. Thus, iSNR becomes a metric describing the reliability of detecting scene detail. The theoretical detection limit is 1, but the reliability requirements of applications may require higher thresholds. For example, a minimum of 10 is suggested for "acceptable" photographic image quality, and between 1.5 and 4 are estimated for machine-vision applications [18].

Technologies to Extend Dynamic Range

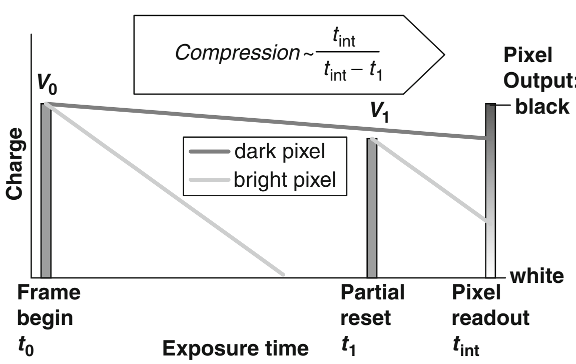
The following section reviews a selection of WDR technologies and explores the iSNR trade-offs of nonlinear dynamic range extension. A WDR video camera for automotive applications should combine good sensitivity with low noise at a given frame rate (10–60 fps) plus a WDR extension that can be adapted quickly to match typical automotive photospace conditions (80–120 dB). The first group of technologies for WDR extension are those that directly change the CMOS pixel response function (such as logarithmic response [21], linear-log combination [22], in-pixel light-to-frequency conversion [23], pixel-level A/D conversion [24], or pixels with overflow capacitors [25]), while a second group changes pixel response indirectly by altering the timing (multiple exposure, multiple partial reset [26]).

The first group, directly changing the response function, involves extensive pixel circuitry resulting in low fill-factor and thus poor low-light sensitivity. For example, in-pixel light-to-frequency conversion provides 115–130 dB WDR, but the required circuitry decreases the fill factor to 25%, and reduces low-light sensitivity. Similar sensitivity degradation currently limits the usefulness for automotive video of both pixel-level A/D conversion and pixels with overflow capacitors. Although pixels with direct logarithmic compression can provide WDR above 170 dB at low cost, the WDR extension is not adaptive, and at present, high fixed-pattern noise and dark current decrease low-light SNR. The combined linear-log CMOS pixel does avoid these drawbacks, but the required seven transistors per pixel limit the fill factor.

WDR can be covered with a linear image sensor by multiple exposure, but application to automotive video requires either fast readout (high frame rate with corresponding loss in low-light sensitivity), or local frame storage and processing.

Among WDR technologies that alter the timing, CMOS image sensors with multiple partial reset (multiple-slope) technology offer good imaging performance at low cost and are thus well suited to automotive video capture [4]. The example of a multiple-slope imager explains how to extend dynamic range while maintaining detection reliability.

At the heart of the CMOS pixel is a light-sensitive photodiode with a parasitic capacitance that is charged by applying a voltage pulse V_0 at frame begin t_0, then discharged as exposure generates photoelectron-hole pairs. Figure 10 shows the charge–time diagram of a CMOS pixel. The rate of discharge depends on the input luminance; a darker pixel is discharged at a lower rate, a brighter at a higher one. After integration time t_{int}, the remaining charge is read out and converted into the image signal D. This is proportional to the degree of discharge between t_0 and t_{int}; no discharge corresponds to $D = 0$, and full discharge to saturation level (Fig. 10). A darker pixel will only be partially discharged at t_{int}, proportional to the input luminance; its response will therefore resemble that of a linear pixel. A very bright pixel will be fully discharged before t_{int}, making it impossible to register any further increase in luminance; it has reached the saturation level of its response curve. However, this bright pixel could resume registering photons if a reset voltage pulse V_1 is applied at the time $t_1 < t_{\text{int}}$ to replenish the charge of each capacitor. The reset is only partial; the reset voltage V_1 is chosen to be lower than the initial V_0 so as to discriminate between darker pixels that would not be saturated at t_{int} and brighter ones that would. The partial reset does not affect those darker pixels that are still above the reset level V_1. It only recharges the brighter pixels that are already discharged below V_1. For them, the reset overwrites any memory of previous exposure, and they resume registering photons in the shorter exposure time $t_{\text{int}} - t_1$. The corresponding pixel response curve will be piecewise linear, but with a lower incremental gain after the breakpoint that corresponds to the reset at t_1. The nearer t_1 is to t_{int}, the lower the incremental gain after reset, and the higher the saturation luminance $L_{\max}$ becomes. The reset allows a higher luminance range to be "compressed" into the same limited output image signal range; thus, the ratio of the total exposure time t_{int} to the remaining exposure time $t_{\text{int}} - t_1$ is referred to as "compression." To increase compression further, additional partial resets of successively decreasing voltage can be applied within the remaining time between t_1 and t_{int}, see Fig. 11a, thus creating a piecewise linear response curve where the incremental gain is further decreased after each of the breakpoints, see Fig. 11b. Theoretically, the number of additional resets is only limited by the smallest increments of timing and recharge voltage control, but in reality, it is limited by the requirement of detection reliability.

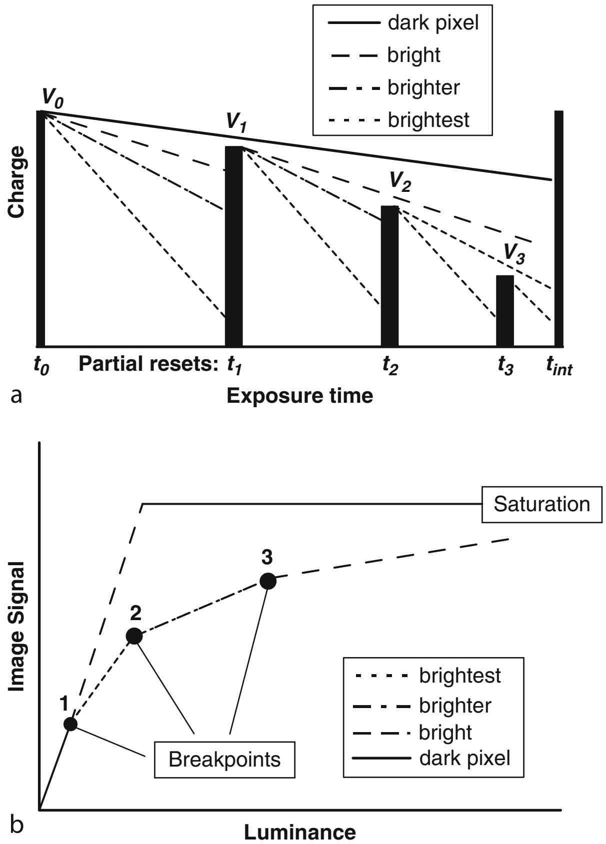


True Color Night Vision Video Systems in Intelligent Vehicles. Figure 10
Charge–time diagram for a CMOS WDR pixel with a single partial reset [5]

Extending Dynamic Range Reliably

The partial resets of a multiple-slope WDR imager require precise control of the timing and recharge voltage so that it can adapt optimally to the changing dynamic range of a natural scene. Experimental observations and model calculations show how the iSNR criterion can be used to maximize dynamic range without compromising detection of scene detail. The methodology for measuring iSNR has been described in detail [5].

Experimental Observations One important goal of dynamic range extension is minimizing the number of saturated pixels and thus maximizing the capture of



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 11
(**a**) Charge–time diagram and (**b**) response curve for a CMOS WDR pixel with multiple partial resets [5]

highlight detail. With the single-reset system shown in Fig. 10, high compression can be achieved by moving the reset time t_1 as close as possible to the frame end t_{int}. The highest compression is limited only by the increments of reset timing, given by one row time in common multiple-slope imagers and by one pixel time for the most advanced multiple-slope imagers. This can be explained in the example of a VGA imager with 480 rows of 640 pixels. The voltage reset can be placed as close as one row time before t_{int} (i.e., after 479 rows), achieving a maximum compression of $c = 480/(480\text{–}479) = 480$. This means that the ratio of exposure times in compressed and linear modes is 1:480 and that the pixel is 480× less sensitive in compressed mode. From exposure time ratios, dynamic range can be estimated to be extendable by 54 dB. In an advanced VGA imager where single pixel times are addressable, the dynamic range extension could be even more impressive. Each row contains 640 pixels. The shortest exposure time would be one pixel time before $t_{\text{int}} = 480 \times 640$ pixels so that a maximum compression of $c = (640 \times 480)/(640\text{–}639) = 307{,}200$ could be reached, extending the dynamic range by 110 dB.

Experiments like those shown in Fig. 12 and the corresponding imager response model of section "Imager Response Model" show that although the practical implementation of high compression with just one partial reset can effectively prevent highlight saturation, it does come at the cost of usefulness, seen as severe loss of midtone image contrast. Figure 12a shows an image taken in linear mode where saturation prevents visibility of detail on the trunk of the car. Progressing from Fig. 12b to d, increased compression decreases the proportion of saturated pixels, making highlight detail more visible on the trunk but at the cost of lowered contrast of midtone detail (lane markings versus road surface). Figure 12d shows the limit of a single-reset system. There are still a fair number of saturated pixels, but the lane markings have become almost indistinguishable from the road surface, which has lost all texture detail. There is a trade-off; the higher the compression, the lower the midtone contrast. At levels of compression still far from theoretical limits, portions of midtone detail almost disappear, and the image becomes unnatural and less useful. The underlying mechanism of this is explained by the following imager response model.

Imager Response Model Response curve and noise models show that the loss of midtone detail as observed in Fig. 12 can be explained by a drop of iSNR that occurs at the reset point. Figure 13 shows modeled curves of OECF and iSNR as functions of log luminance. The model assumes a constant dark noise floor without considering photonic shot noise.

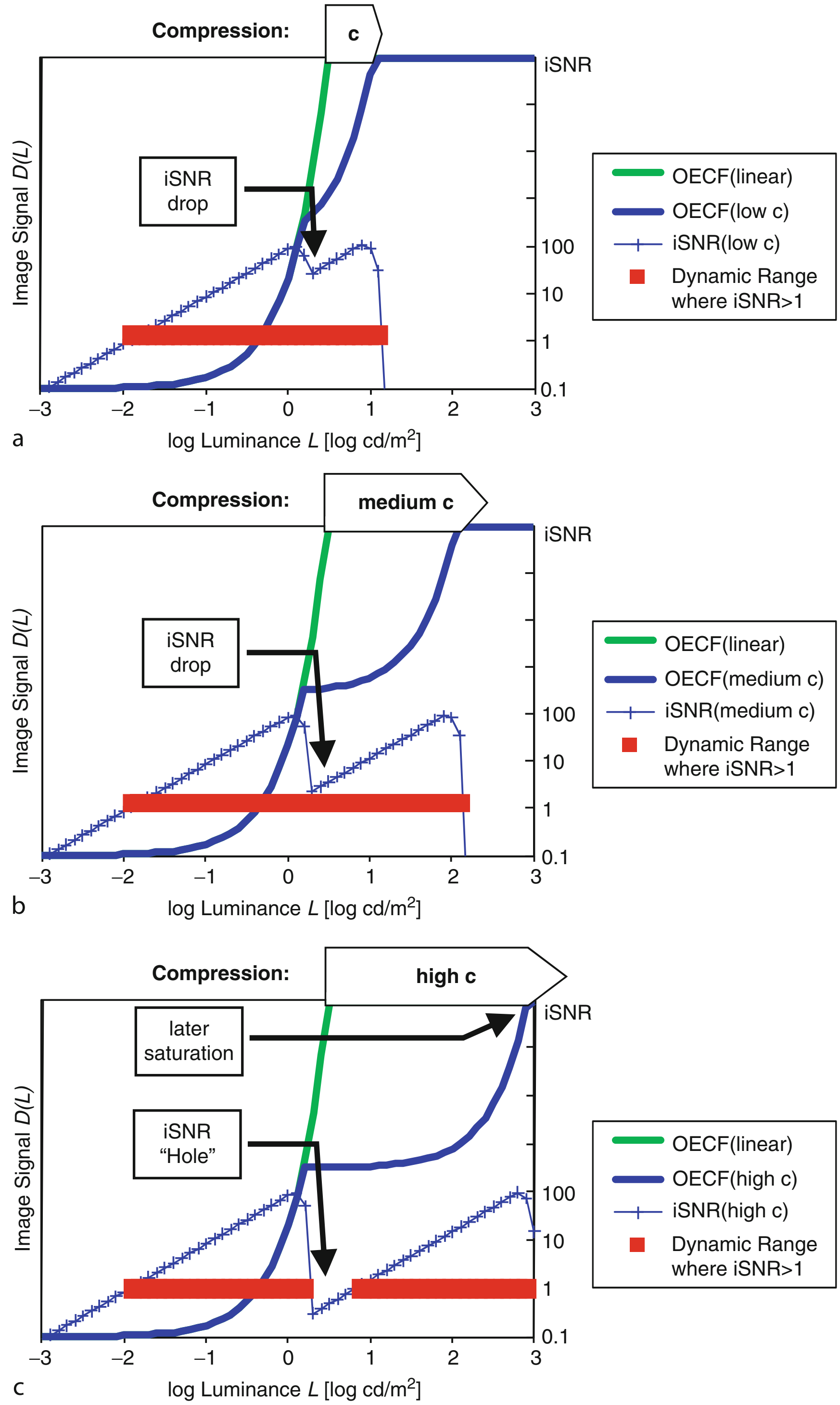
After the partial reset breakpoint, the slope of the response curve (and thus the incremental signal) is reduced so that saturation is delayed to a correspondingly higher luminance level. With increased compression, the drop in iSNR at the breakpoint becomes deeper until it drops below the reliability threshold as shown in Fig. 13c, thus creating an undesirable zone of extinguished local contrast, a so-called iSNR hole. The luminance range

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 12
Natural scene captured with (**a**) linear camera and with a single-reset WDR camera at (**b**) low, (**c**) medium, and (**d**) high compression, demonstrating the trade-off between increased dynamic range and decreased midtone contrast [5]

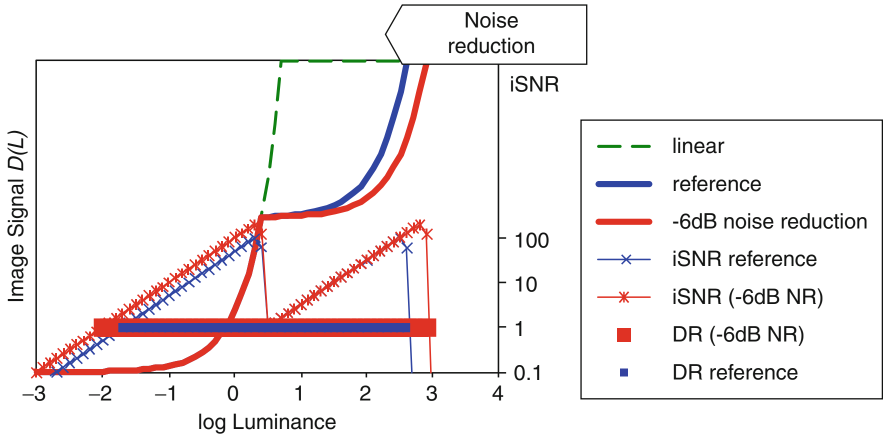
within and near the "iSNR hole" corresponds to the luminance levels of those details such as road texture and lane markings that became almost indistinguishable in Fig. 12d.

Along with saturated pixels or those below the noise floor, "iSNR holes" make the detection of scene detail impossible. In linear imagers, the zones of iSNR failure occur only at very dark pixels (below noise floor) and very bright pixels (above saturation level). However, in multiple-slope WDR imagers, zones of iSNR failure can also occur at luminance levels corresponding to reset times. These levels cannot be predicted without exact knowledge of the response curve for a given set of reset times. If the resets are automatically adapted to the intra-scene dynamic range, then the luminance levels corresponding to the reset times would depend on scene content and thus be variable, making iSNR holes potentially interruptive to the visibility and automatic detection of scene detail. It follows that resets have to be controlled to prevent iSNR holes occurring in the first place. Dynamic range can only be extended as long as the incremental signal *iS* can be kept above the noise floor at every luminance level. Thus, iSNR is not just a criterion for determining the upper and lower dynamic range limits but becomes critical when controlling the reset times within the dynamic range. Adaptive WDR extension requires separate control loops for integration time and compression that both adapt the imager's response curve to include the darkest and brightest scene details. The control algorithms for timing the multiple partial resets in adaptive WDR extension must be designed specifically to avoid iSNR holes.

Reliably Extending Dynamic Range The example of a multiple-slope WDR camera demonstrates how the iSNR criterion can be used to extend dynamic range without sacrificing detection reliability.



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 13
Log-linear response curves (OECF) and incremental SNR of a single-reset model at (**a**) low compression, (**b**) medium, and (**c**) high compression [5]



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 14
Lower pixel-level noise increases dynamic range in a single-reset imager [5]

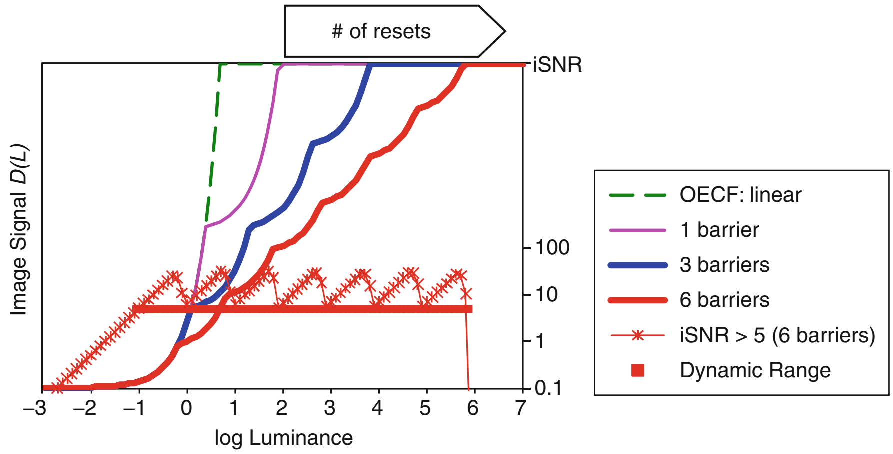
If the number of resets is limited to one or two, then noise reduction is the most effective way of increasing dynamic range. Figure 14 compares two single-reset model imagers, one where altered pixel design (i.e., larger size, correlated double sampling, etc.) has cut noise in half (-6 dB). This leads to a twofold increase in dynamic range. Firstly, the 6 dB lower noise floor leads to a 6 dB lower noise-equivalent luminance signal $L_{\min}$. In addition, the lower noise floor allows – without creating an iSNR hole – the compression to be increased by a factor of 2, halving the slope after reset and doubling the saturation luminance $L_{\max}$, thus increasing the dynamic range by another 6 dB, a total of 12 dB. Assuming a minimum iSNR of 1, dynamic range is increased from 88 to 100 dB or twice that of a linear imager of 50 dB.

Where pixel-level noise reduction is not effective (small pixels, low fill factor, high operating temperatures, inexpensive pixel design), then increasing the number of partial resets is a better strategy. The drop in iSNR at a reset point depends on the ratio of incremental gains after and before the reset, $k = g/g_0$. The more $g(L)$ is reduced by a reset, the deeper iSNR will drop. After each partial reset n, the sensitivity decreases geometrically by the factor $1/k^n$. When deploying multiple partial resets, their timing and charge (voltage) can be chosen so that the linear output signal range is equally distributed between the resets [5]. This ensures that iSNR at each reset point drops to the same minimum level (see Fig. 15). Having more resets has the added advantages: the decrease in incremental gain can be approached more gradually, and the iSNR threshold can be raised from one to a more practical level. Figure 15 shows a dynamic range model for an increasing number of resets. Assuming an increased iSNR threshold of 5, the linear imager will have a dynamic range of only 36 dB. Compression with a single reset will extend the dynamic range to 62 dB; two resets will reach 98 dB, and six resets 138 dB.

Conclusion

Although the concept of iSNR-based reliability has been modeled and tested only for multiple-slope CMOS imagers, it can be deduced that the concept applies to any other WDR extension technology that involves sudden changes in the incremental gain of the pixel response curve. In contrast, sensors with logarithmic response are in no danger of "iSNR holes" as their incremental gain decreases continuously during exposure time.

For the example of multiple-slope CMOS imagers, it has been shown that iSNR is a critical criterion governing the number, timing, and recharge voltage of partial resets. Dynamic range can be increased by reducing noise at the pixel level, by increasing the number of partial resets, and by distributing them equally to maintain a minimum iSNR necessary for reliable detail detection throughout the entire dynamic range. As technologies for extending the dynamic range evolve, new solutions will be added to the repertoire and tested for their suitability for intelligent vehicle



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 15
More resets increase dynamic range while allowing a higher minimum iSNR threshold of 5 [5]

technology. However, the methodology described here will provide a guideline for optimizing the pixel response for maximum dynamic range at the highest possible iSNR.

Extending Spectral Sensitivity into NIR

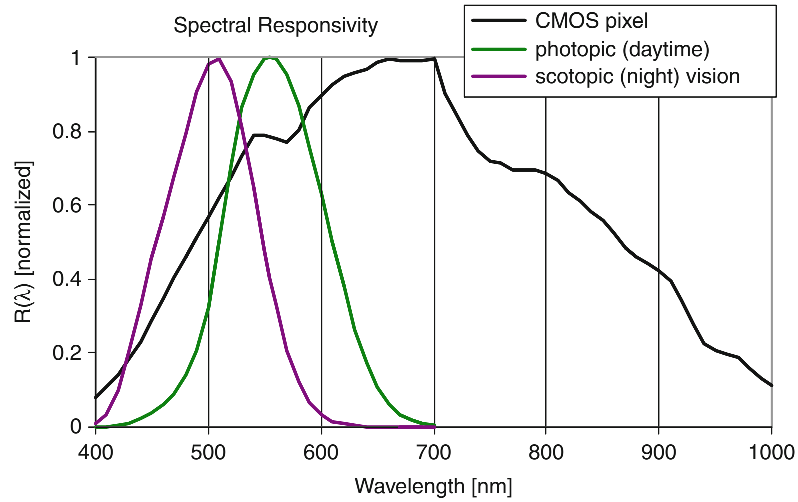
In order to capture dark scene detail illuminated by ambient light, the camera has to be very sensitive with high SNR. Exposure time is limited by the video frame rate and the need to limit motion blur. Sensitivity can be increased by either lowering noise (increasing SNR) and/or extending the range of spectral energy sensed.

IR Sensing

The spectral sensitivity of silicon-based image sensors extends beyond the long-wavelength edge of the visible spectrum (730 nm) into the infrared (IR). As longer wavelengths penetrate further into silicon, CMOS imagers can be made more sensitive to IR by collecting photo charge carriers from greater depth and by increasing substrate thickness. The IR cut-off in CMOS imagers can thus be shifted as high as 1,100 nm. This spectral range coincides with the shorter wavelengths of the IR-A or near-infrared (NIR) range (700–1,400 nm). Imagers for longer wavelengths beyond the NIR rely on semiconductors with a narrower band gap, for example InGaAs (950–2,600 nm). Such far-infrared (FIR) cameras can sense the black body radiation that surfaces emit as a function of their temperature, and obtain a thermal image of the environment. They do not require any other source of radiation and can operate in complete darkness. Compared to NIR systems, FIR night vision systems are more costly, cannot be placed behind glass, and deliver unnatural images that are difficult to interpret [27, 28].

Passive and Active NIR Vision

The detection of objects by a NIR-sensitive camera requires them to be irradiated by a NIR source so that the camera can detect radiance image from reflected NIR. Passive sensing relies on ambient NIR. Ambient NIR at night comes mostly from artificial incandescent light sources such as vehicle headlights; vapor street lighting and natural light sources (moon, stars) emit comparatively little NIR. Daylight especially direct sunlight contains extremely high background NIR that is strongly reflected by vegetation; cameras may have to operate in WDR mode to avoid signal saturation. Active sensing requires sources that are mounted to the same vehicle as the NIR camera. Conventional metal halide (tungsten) headlamps emit more than 60% of their spectral energy in the NIR. Dedicated NIR-only headlamps, either tungsten lights with a VIS-cut filter or NIR LEDs at 850 nm, have the advantage that they are invisible to humans. As they cannot blind oncoming drivers, they can be operated on high-beam, thus extending the range of vision beyond that of dipped headlights [29]. One of the



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 16
Comparison of photopic and scotopic luminous sensitivity with the spectral sensitivity of a monochrome automotive CMOS image sensor [42]

first active systems implemented was the Mercedes-Benz Night View Assist (2005).

Comparing the spectral sensitivity of a monochrome CMOS image sensor with the photopic and scotopic luminous sensitivity functions shows the range of spectral energy that the image sensor can utilize. For example, when the scene is illuminated by metal halide headlamps whose emission spectra peak in the NIR at about 900 nm, sensitivity stretches to the long-wavelength cut-off of the image sensor (Fig. 16).

The example in Fig. 17 shows an urban traffic scene taken without and with ambient NIR, showing the main benefit of extended NIR sensitivity: better visibility of the roadway where it is illuminated by dipped headlights.

The range of visibility can be extended even further, without blinding oncoming drivers, if dedicated NIR-only high-beam headlights are used. These will, of course, "blind" any NIR-sensitive night vision cameras of other vehicles, unless they are capable of performing in a WDR environment.

Adding Color Differentiation to VIS + NIR-Sensitive Imager

Using NIR to increase the sensitivity of color imagers requires separation of the NIR and RGB color signals because the color filter materials used in standard CMOS and CCD technology are highly transparent to NIR so that each color pixel also detects NIR (Fig. 18).

Mixing color and NIR signals can result in extreme color de-saturation if the illumination contains high amounts of IR. This is the case when incandescent lamps are used with color temperatures typically about 3,200 K, where spectral power distribution peaks at about 800 nm (Fig. 7). Here, the signal contributions to RGB from the NIR are even higher than those from the actual colors so that all colors are rendered as off-gray, for example shades of pink (Fig. 19). The example demonstrates that the differentiation of color requires the NIR signal to be separated from the color signals.

Increasing the Spectral Sensitivity of Color Imagers

Color image sensors simply suppress the NIR information by an IR-cut filter. The sensitivity of such imagers can be improved by adding filterless "white" pixels to the RGB filter pattern [30, 31] (Fig. 20). Compared to the RGB color pixels, these "white" pixels have about three times higher spectral sensitivity. As scene illuminance decreases, information from color pixels is increasingly mixed with luminance information from the W-pixels by pixel binning. This requires high pixel counts plus special (nonstandard)

a

b

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 17
Monochrome image of urban traffic scene (**a**) without NIR and (**b**) with NIR

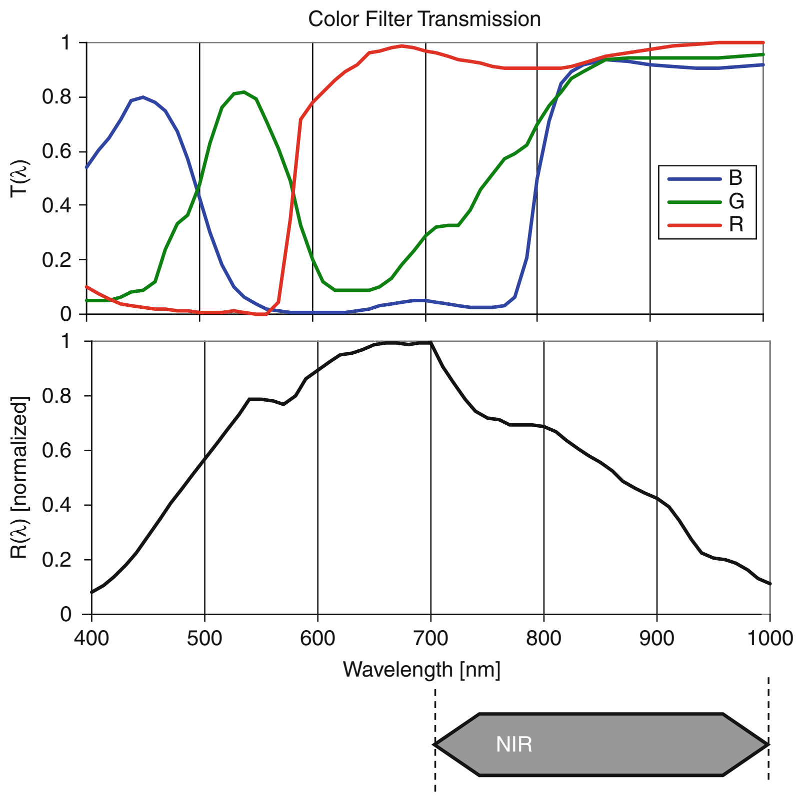
filter patterns, and the NIR remains unutilized for night vision.

Extending the Spectral Sensitivity of Color Image Sensors into the NIR

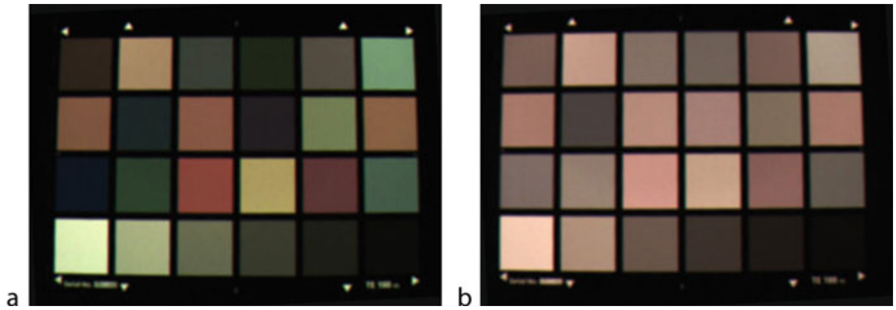
Extending the light sensitivity of color imager sensors into the NIR requires a separation of the NIR and the color signals. This could be achieved by using a beam splitter and two image sensors, one monochrome for NIR, and one color for VIS. However, cost and form factor do not recommend this solution for automotive vision applications.

Single-imager solutions are based on the addition of a fourth pixel type to the RGB-pixels (Fig. 20c). This fourth pixel, which, for example, replaces the second G-pixel in the Bayer pattern, is either NIR-sensitive or "panchromatic," sensitive to both visible and NIR light. A variety of methods have been proposed to separate NIR signals from color signals:

- Patterned optical NIR-cut filters can be added to mosaic filter arrays, for example RGBW or CymG (cyan-yellow-magenta-green) [32], to protect all or a portion of color pixels from the effects of NIR radiation. The filters can be of the absorptive or interference type. Interference filters are preferred for their sharper absorption edge toward the NIR, but they do require patterned deposition of more than ten dielectric layers [36], which is both costly compared to dye-based filters and uses dielectric materials that are incompatible with standard Si processes. The effectiveness of such filters is limited by optical crosstalk of NIR into the color pixels after passing the NIR cut filter, especially at low incident angles.
- Nano-optical filters, currently still in the research stage, can be designed as short-pass edge filters or band-pass filters. They can either be integrated as metal layers into the pixel stack [37] or deposited



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 18
VIS–NIR spectral transmission curves of typical RGB color filters used in CMOS imagers [35]

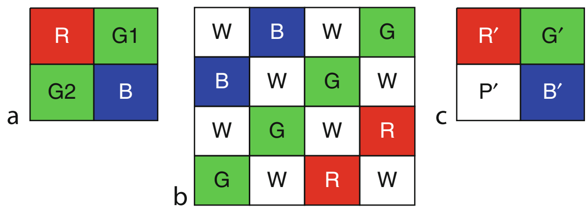


True Color Night Vision Video Systems in Intelligent Vehicles. Figure 19
Raw images of a Macbeth color checker chart illuminated by incandescent light (3,200 K) taken with a CMOS color camera (**a**) with IR-cut filter and (**b**) without IR-cut filter [35]

on top of the imager [38]. Although nanostructured metal layers within the pixel stack have to follow the Si design rules for structure size (0.15, 0.13, 0.11, 0.09 μm) which limits the degrees of freedom for spectral tuning, they are practically free from optical crosstalk. Filters on top of the imager can be created by nanoimprint technologies which allow the creation of an array of plasmonic band-pass filters. However, the peak transmission of band-pass filters is relatively

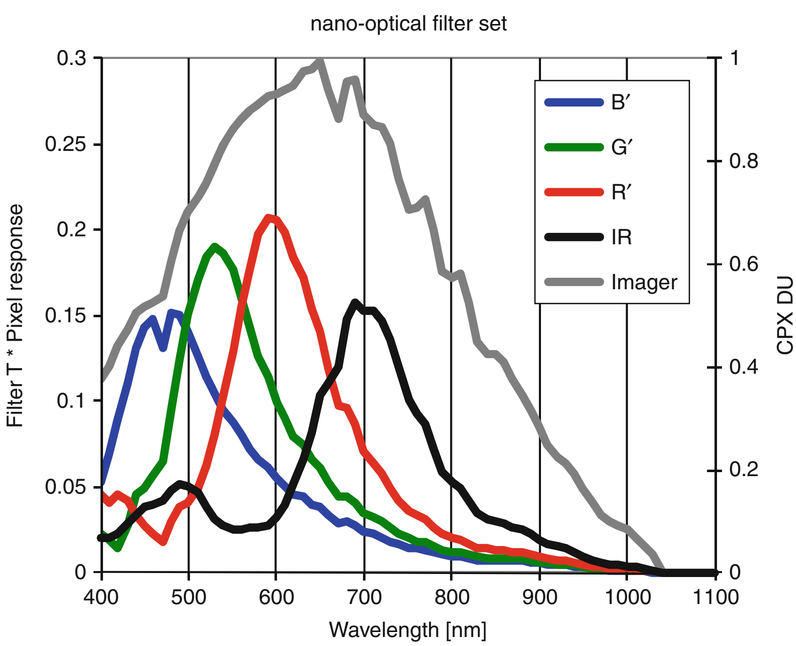
low (<0.4). Peak filter transmission decreases, and halfwidth increases with wavelength, making nano-optical filters less effective in the NIR (Fig. 21) [38].

- Virtual NIR filters separate NIR and color signals by image signal processing. One possible implementation adds filterless pixels to the RGB filter pattern, but in contrast to RGBW, no optical NIR-cut filter is used [33–35]. The filterless pixel has two functions: Firstly, it has a "panchromatic" spectral response that includes the NIR and is equivalent to that of a monochrome imager. The panchromatic P′-pixel is more than three times as sensitive as a color imager, capable of delivering a much brighter luminance image. Secondly, the P′-pixel delivers the reference signal necessary to estimate the NIR signal contribution to the R′G′B′ pixel signals. The raw color pixel responses are $R' = R + I$, $G' = G + I$, and $B' = B + I$ (Fig. 22). The panchromatic pixel response $P' = R + G + B + I$ provides the reference signal for estimating the NIR portion in the incident light, which is then used in the virtual NIR filter matrix to remove as much as possible of the NIR contributions to the R′G′B′ signals. The response of the P′-pixel can also be used to simply retrieve a panchromatic VIS–NIR luminance channel $L_{VIS-NIR}$.

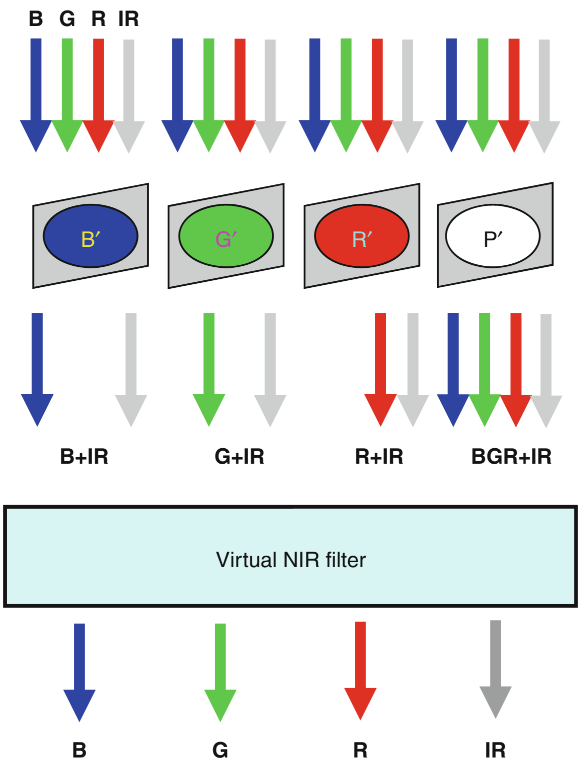


True Color Night Vision Video Systems in Intelligent Vehicles. Figure 20
CFA layouts: (**a**) standard Bayer, (**b**) RGBW, and (**c**) RGBP [35]

An example of raw spectral pixel responses ($R'(\lambda)$, $G'(\lambda)$, $B'(\lambda)$, $P'(\lambda)$) shows the high NIR signal contribution to the color signals as well as high amounts of crosstalk between the four pixel types. Since both deteriorate colors, they have to be minimized by determining



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 21
Simulated response curves for CMOS imager with plasmonic nano-optical filters [38]



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 22
Light, pixel responses of the R′G′B′P′ imager and image signals after virtual NIR filtering [35]

the virtual NIR filter coefficients, for which two calibration methods are commonly used, global and spectral.

- The global method relies on a set of known reference colors, for example those of a Macbeth color checker chart, that have to be reproduced as faithfully as possible. The color match can be done in a standardized color space, for example CIE 1931 *XYZ*, and the matrix M_G optimized so that the transformed colors match the reference colors $[XYZ]$ as closely as possible,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \approx M_G \cdot \begin{bmatrix} R' \\ G' \\ B' \\ P' \end{bmatrix}. \tag{5}$$

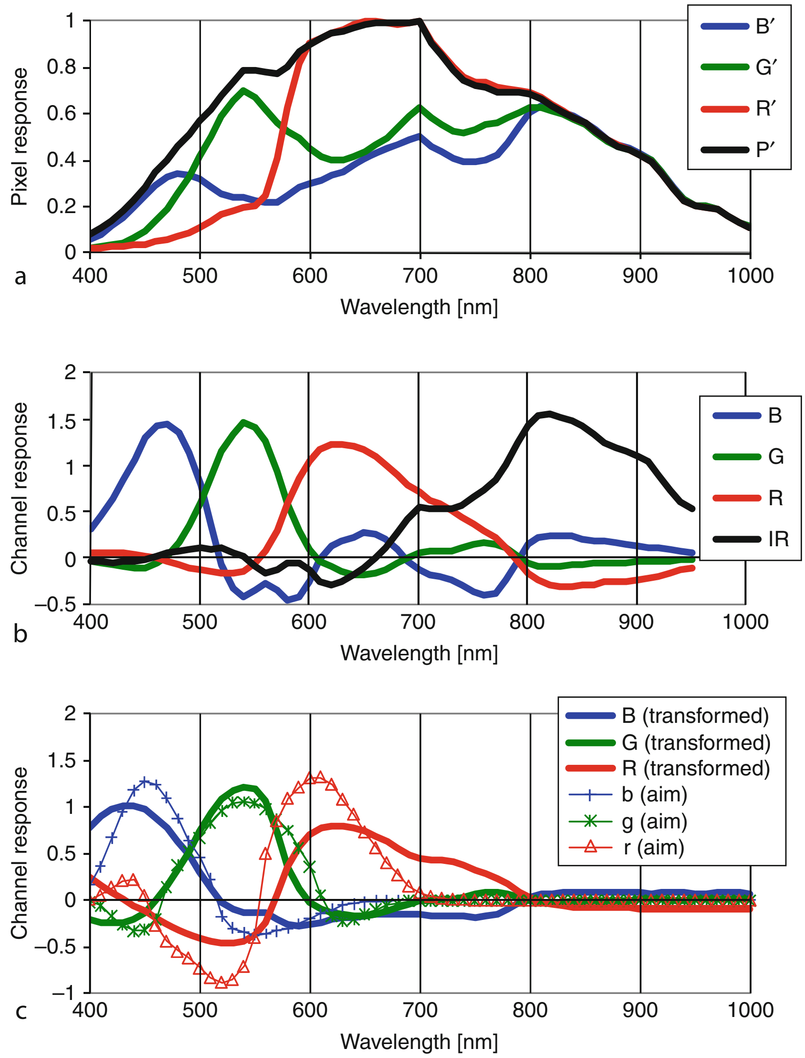
  Global calibration works well for colors within and close to the reference set and under illumination conditions similar to those used in the calibration, but it is poorly suited for color night vision where the most relevant colors are not reflective but rather emissive colors whose brightness might lie far outside the set, and whose illuminant, and thus, relative NIR content varies widely.
- The spectral calibration method optimizes the matrix coefficients β_{mn} so that the filtered spectral response functions $(R(\lambda), G(\lambda), B(\lambda), I(\lambda))$ fulfill application-specific conditions.

$$\begin{bmatrix} R(\lambda) \\ G(\lambda) \\ B(\lambda) \\ I(\lambda) \end{bmatrix} = \begin{bmatrix} \beta_{11} & -\beta_{12} & -\beta_{13} & \beta_{14} \\ -\beta_{21} & \beta_{22} & -\beta_{23} & \beta_{24} \\ -\beta_{31} & -\beta_{32} & \beta_{33} & \beta_{34} \\ \beta_{41} & \beta_{42} & \beta_{43} & -\beta_{44} \end{bmatrix} \cdot \begin{bmatrix} R'(\lambda) \\ G'(\lambda) \\ B'(\lambda) \\ P'(\lambda) \end{bmatrix} \tag{6}$$

For example, in a machine-vision application, the coefficients are optimized to remove from the transformed spectral responses $(R(\lambda), G(\lambda), B(\lambda), I(\lambda))$ as much of the NIR and color crosstalk signals as possible. Instead of retrieving an IR channel, the P′- response can be mapped into a panchromatic luminance channel $L_{VIS-NIR}$. In human vision applications, a colorimetric calibration is better suited for achieving natural-looking colors. Colorimetric calibration as shown in Fig. 23c tries to optimize the matrix M_S so that the filtered spectral response curves of the imager approximate the spectral color matching functions $(b(\lambda), g(\lambda), r(\lambda))$ of the human visual system as specified in ISO 17321 [39]:

$$\begin{bmatrix} r(\lambda) \\ g(\lambda) \\ b(\lambda) \end{bmatrix} \approx M_S \cdot \begin{bmatrix} R'(\lambda) \\ G'(\lambda) \\ B'(\lambda) \\ P'(\lambda) \end{bmatrix}. \tag{7}$$

Spectral calibrations have the advantage of not relying on a particular set of reference colors, and also, they do not rely on a specific illuminant. The virtual NIR filter with colorimetric spectral calibration reduces the NIR crosstalk into color signals even under illumination conditions with high amounts of NIR, for example incandescent light at 3,200 K color temperature. The example in Fig. 24 shows images before (a), (b), and after virtual NIR filtering (c) with increasing amounts of NIR. Increasing NIR reduces color saturation of the raw R′G′B′ color channels, and increases brightness of the panchromatic P′-channel. The colors reconstructed by the virtual NIR filters become darker



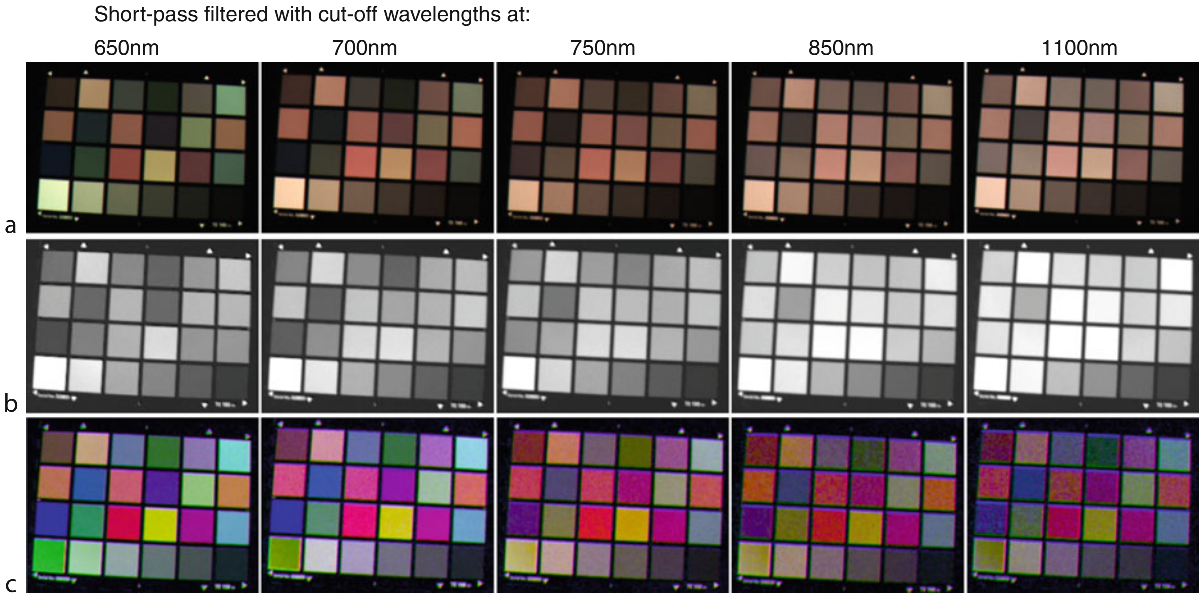
True Color Night Vision Video Systems in Intelligent Vehicles. Figure 23
Spectral pixel responses of (**a**) the R′, G′, B′, and P′ pixels and spectral channels after (**b**) separating NIR from the RGB color channels and (**c**) colorimetric calibration [42]

with increasing levels of NIR because of decreasing color-to-NIR signal ratios in the raw images.

However, residual spectral leakage limits the performance of the virtual NIR filter, especially in WDR mode where initial signal ratios are reduced. Careful management of NIR can further reduce the effects of NIR crosstalk and has to include pixel and microlens design to minimize optical and electrical crosstalk, high SNR to counter the noise-amplifying effect of high filter coefficients, and NIR-only headlights to minimize the proportion of NIR radiation with the highest leakage into the color signals.

Extending the Dynamic Range of Color Imagers with NIR Sensitivity

Signal saturation in color imagers leads to loss in color information; as the signals in the R, G, and B channels approach saturation levels, colorfulness decreases. In a RGB color imager, signal clipping in one or more color channels leads to a whitening of colors; clipping in a four-channel system such as R′G′B′P′ can also lead to false colors. WDR extension is therefore essential to capture color detail carried by bright lights. However, corrections have to be made for the effects of WDR



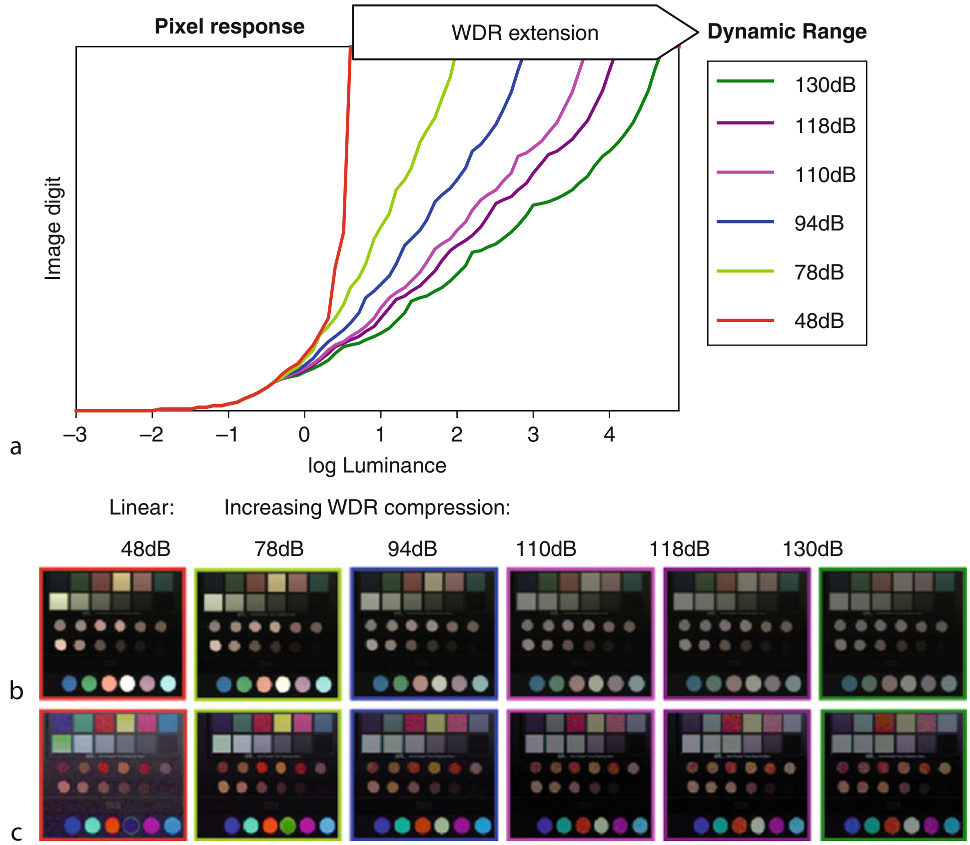
True Color Night Vision Video Systems in Intelligent Vehicles. Figure 24
Macbeth color checker chart illuminated by incandescent light (3,200 K) and imaged with short-pass filters allowing increasing amounts of NIR: (**a**) raw R′G′B′ image, (**b**) raw P′ image, and (**c**) RGB image after virtual NIR filtering [42]. The false colors in the brightest patch are caused by clipping of the P′-signal

extension on the color signals. In the example of multiple exposures with varying exposure time, each exposure maps a different range of input radiance signals into the same range of output image signals. Within the sequence of images to be combined into a single WDR image, the same output signal may be produced by different combinations of input signals so that the relationship between input and output signals is not unique. Therefore, the combination of exposures into the WDR image has to be preceded by linearization, which uses the camera response function in each color channel to create a unique map between the digital signal levels in the WDR image and the input radiance levels. The response function can be derived by techniques such as those proposed by Debevec and Malik [40], or Mitsunga and Nayar [41].

Direct WDR methods that extend the dynamic range in a single exposure have the great advantage of providing a unique map between input radiance and output image signal levels (Fig. 9). However, if the response function is nonlinear, for example piecewise linear as shown in Fig. 11, the signal ratios between the color channels will decrease with decreasing slope of the response curve. Increasing compression will thus decrease the input signal ratios between the raw R′G′B′P′ channels, thus decreasing chromaticity and iSNR of the output colors after virtual NIR filtering. To recover the color components and remove the NIR component from the highly compressed signals of bright colored lights, the raw R′G′B′P′ pixel signals have to be linearized with the inverse response curve of the imager so that the image signals can be referred back to the corresponding scene radiances (inverse gamma correction). The effect of linearization is shown in Fig. 25 where it restores the colors in the highly compressed images of bright colored lights below the Macbeth color checker chart [42].

Visual Representation of the VIS + NIR Color Information

Section “Conclusion: Producing Natural Images Under Mesopic Conditions” outlined the requirements for rendering a natural-looking night vision image. In the example of a WDR imager with panchromatic and color pixels, NIR sensitivity increases the range of



True Color Night Vision Video Systems in Intelligent Vehicles. Figure 25
WDR color scene consisting of dark reflective colors (*top*), bright transmissive colors with (*center*) and without NIR (*bottom*), taken with (**a**) imager response of increased compression and displayed (**b**) without and (**c**) with linearization [42]

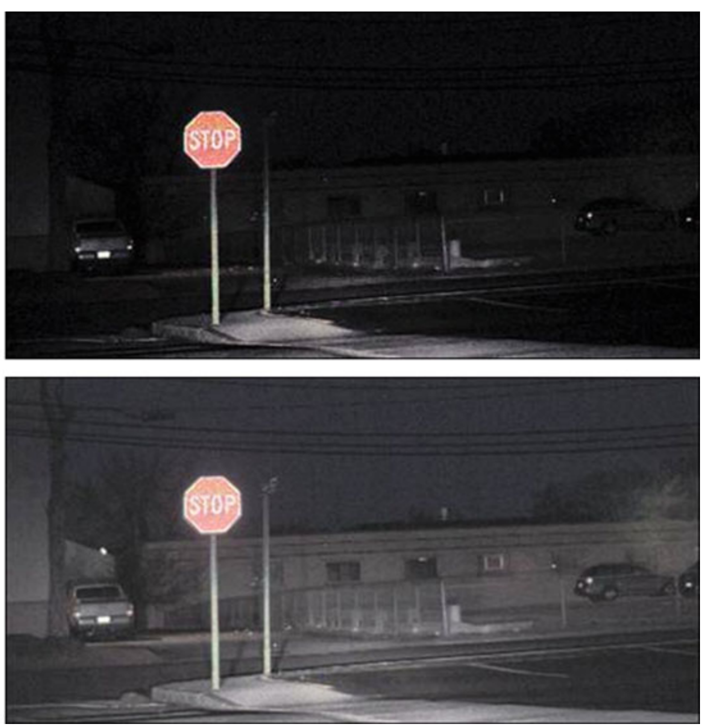
visibility, and the panchromatic image information can be displayed in the luminance channel. The virtual NIR-cut filter can recover traffic-relevant color information from backlit or retroreflective signage, traffic lights, and taillights. Virtual NIR filtering shows reasonably good performance in reconstructing colors even under illumination conditions with high amounts of NIR, for example incandescent light at a color temperature of 3,200 K having 20% signal in the visual and 80% in the NIR range (Fig. 24). Since the filter array relies on standard RGB filter materials and fabrication processes, it can be produced at comparably low costs as conventional RGB imagers, without the extra cost of a patterned optical NIR-cut filter.

The VIS and NIR image information has to be represented in a way that the human observer viewing video on a dashboard screen can easily interpret. The VIS + NIR color camera with virtual NIR filter can provide multispectral image information, for example, in four channels, RGB and NIR. Multichannel image data could simply be mapped into a false-color representation, but this creates unnatural images that do not agree well with visual expectations and can even be confusing rather than useful. Natural image data representation that agrees with human expectations can be achieved by following certain design guidelines [43]: (1) orthogonal encoding to represent information with as little redundancy as possible; (2) anthropometric optimality to maximize usefulness to the human observer; (3) ecological invariance so ensure similarity of image interpretation over a variety of environmental parameters such as time of day, illuminants, and weather conditions; (4) computational simplicity to allow video implementation. For example, the achromatic channel of the HVS carries over

95% of the scene information and is much more sensitive to spatial information than the opponent color channels which process color information. Therefore, true-color image fusion is best suited for this application because it uses the extra information from the NIR to enhance the spatial and luminance information in the achromatic channel while mapping the color information from the visible range into the chrominance channels [32, 44, 45]. First, the RGB image obtained by virtual NIR filtering (6) must be transformed into a luminance–chrominance color space; then, its luminance component L_{VIS} is replaced by the luminance $L_{VIS+NIR}$ estimated from the P′-pixels. Additional processing of $L_{VIS+NIR}$ ensures that the fusion increases the brightness of the dark scene background but not so much that of colored lights, thus preventing the loss of color saturation in colored lights after fusion. For example, tone mapping using global, local, or Retinex algorithms significantly improves the visibility of grayscale and color detail in WDR images [46].

The night scene in Fig. 26, illuminated only by the car's headlights, shows how image fusion effectively extends the field of vision, increases visibility of detail, and retains color in the retroreflective sign.

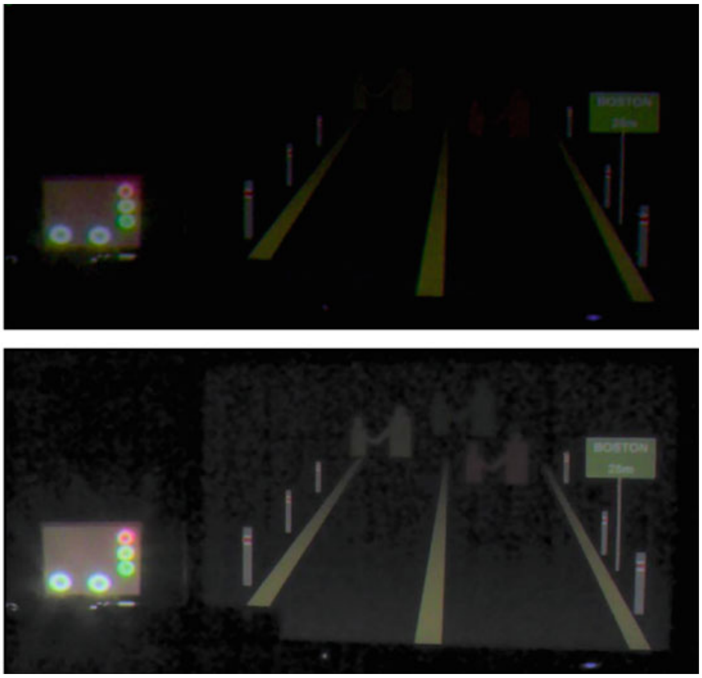
A previously described WDR scene simulator [5] composed of a digital display and a projector has been used for creating test scenes with controlled luminance levels, covering a wide dynamic range of 100 dB. Individual scene elements such as the road surface, lane markings, traffic signs, pedestrians, and headlights shown in Fig. 27 were calibrated so that their absolute luminance levels were within the ranges of typical traffic scenes as shown in Fig. 3. The example demonstrates that even if little or no IR is present in the scene, image fusion will still increase the brightness of dark scene detail because the filterless P′-pixels are more sensitive than any of the color pixels. Fusion utilizes the brighter luminance image from the P′-pixels, making the pedestrians in the dark background much apparent. The retroreflective sign and traffic lights are also brightened, but only to the extent that color saturation is not adversely affected.



a

b

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 26
Night scene with tungsten headlights as only light source, taken without image fusion (**a**) and with image fusion (**b**)



a

b

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 27
Simulated 100 dB WDR traffic color scene (**a**) without and (**b**) with image fusion [35]

In real-world night vision scenes, the panchromatic pixel can utilize the large NIR component of incandescent (tungsten) headlights that is invisible to the driver (Fig. 7). Furthermore, high-beam NIR-only headlights can be used to increase visibility beyond the distance range of headlights at low beam.

Example: Render Traffic-Relevant Color Information in Natural Images

Figure 28 demonstrates the image rendering of a VIS–NIR true-color WDR night vision camera for a typical urban night traffic scene.

The monochrome image in Fig. 28a shows that the wide spectral sensitivity of the panchromatic P′-pixels delivers a sufficiently bright monochrome image even at a frame rate of 30 fps. The color image in Fig. 28b is comparatively dark, consisting mainly of the "unrelated colors" of traffic lights and vehicle taillights against an almost black background. The spectral energy in the visible band is small compared to the total energy within the imager's sensitivity range so that the luminance component of the filtered colors is much lower than the luminance from the P′-pixels. WDR compression and subsequent linearization ensure that the bright highlight colors are preserved. The fused image in Fig. 28c then combines the bright monochrome background with the unrelated colors, leading to a more natural rendering of the scene that contains all traffic-relevant color information from traffic lights and car signal lights.

Conclusion

True-color night vision systems that utilize the NIR sensitivity and WDR capability of CMOS image sensors can aid the driver in dangerous situations where human vision is inadequate or challenged (Table 2). NIR sensitivity extends the range of vision when the safe braking distance exceeds the range of headlights. Cameras with adaptive WDR compression provide natural images of scenes where the intra-scene dynamic range exceeds that of the human eye, for instance when blinded by headlights of oncoming vehicles. Such cameras are able to

True Color Night Vision Video Systems in Intelligent Vehicles. Figure 28
Urban night traffic scene: (**a**) panchromatic P′ image, (**b**) RGB color image after virtual NIR filtering, and (**c**) final image after image fusion [8]

adapt to sudden changes in scene brightness much faster than the HVS so that image information is available while the eye is still in the process of adapting, for example to the "black hole" after bright headlights of a passing car have left the field of vision.

Future Directions

Vision-based DAS are one of a wide range of intelligent vehicle technologies that can be implemented either as stand-alone or cooperative systems. Some are already in use (antilock braking system, electronic stability control) while others are still in the research stage,

True Color Night Vision Video Systems in Intelligent Vehicles. Table 2 Advantages of camera-based night vision systems

	Human visual system	True-color VIS–NIR camera with WDR	Advantages of camera
NIR sensitivity	No	Up to 1,100 nm	Extended range of vision
Adaptation to brightness and dynamic range	Transient 1–2 s	3–6 frames (0.2–0.1 s)	Image information available at WDR and during sudden changes in brightness
Color vision	Mesopic shift	Color constancy	Consistent color information

under development, or in the process of being introduced into the market [47, 48].

Vision-based driver assistance systems aim to enhance the visual perception of the road ahead in situations where the human visual system is challenged or even fails to perceive dangerous situations. The scene information captured by the camera is processed into a natural-looking image and displayed in the driver's field of view. Since more than 90% of driving decisions and actions are based on vision, the fastest and most efficient human–machine interface is a visual display. Their combination of relative simplicity with high capacity and the speed of the visual information channel recommended display-based vision systems to early proponents of intelligent vehicle technology. Active NIR-sensitive night vision systems such as the Toyota Night View (2002) and Mercedes-Benz Night View Assist (2005) were pioneered in top-of-the-range cars. Despite the relative simplicity of camera-to-display night vision systems versus machine-vision-based automatic driver assistance systems (such as lane keeping, obstacle detection, etc.), the technical challenges and reliability requirements are so high that their penetration into the market is still low. The night vision system must deliver useful natural images under an extremely wide range of lighting conditions while experiencing changing proportions of color and NIR signals. Furthermore, they have to perform under adverse weather conditions that severely tax human vision, such as rain, drizzle, fog, snow, and salt spray. The performance and reliability of future night vision systems might be improved in many ways, perhaps including:

- Improved NIR management by integrating adaptive NIR-only headlights into the night vision system [49]
- Increased SNR of the image sensor to increase dark sensitivity and counter the increased visibility of color noise after virtual NIR filtering
- Reduced NIR crosstalk into the color signal through CFA and color pixel design, for example, nano-optical color filters [37, 38]
- Reduced NIR leakage in virtual NIR filtering by using algorithms that further reduce NIR leakage, for example, neural networks [50]
- Improved WDR color processing, for example, Retinex-based tone mapping [46]

In addition, head-up displays such as windscreen projection will be able to communicate visual information to the driver without shifting attention away from the road as necessitated by a head-down display outside the field of vision, for example, the dashboard displays of early night vision systems [28].

The main shortcoming of any display-based vision system is that it only improves traffic safety if the driver pays attention to the display at the critical moment. For example, it is of no use if the driver does not check the display as an obstacle beyond the field of vision is displayed by the NIR-sensitive system before the driver can see it through the windshield. Recent developments in intelligent vehicle technology work toward addressing this shortcoming and move beyond the early display-based vision systems. The two main strategies are intelligent stand-alone systems to enhance vehicle safety, and cooperative systems to improve road traffic safety [48].

- Stand-alone, in-vehicle systems use sensors to detect possible dangers and try to avoid or mitigate accidents by following a three tier strategy: (1) inform the driver as early as possible, (2) issue warnings if there is no driver reaction to the information, and (3) actively assist or ultimately

intervene in the vehicle control [48]. In-vehicle systems will rely on multiple vision sensors and combine different sensor technologies such as radar and LIDAR (light detection and ranging) alongside vision sensors to improve overall detection reliability. Display-based vision systems can be expected to remain an important integral part of on-board safety systems because a visual display more immediately informs the driver of the nature of the danger than an acoustic warning would.
- Cooperative systems increase the capacity for gathering information by using communications from infrastructure or other sensor-equipped vehicles to make an individual vehicle less dependent on its own sensors and local environmental conditions. For example, if a car drives into a bank of fog, it can be guided by infrastructure and information from other vehicles outside the danger zone.

Forward-looking VIS–NIR cameras with color discrimination and WDR will continue to produce useful image information for both human and machine vision systems. Even if camera-based night vision systems with visual human–machine interfaces turn out to be only a transition stage toward future intelligent vehicle solutions, the technical experience gleaned by the implementation and application of pioneering vision systems may still prove indispensible for the future development of intelligent vehicle technology.

Bibliography

Primary Literature

1. Traffic Accident Causation in Europe (TRACE) (2007) Review of crash effectiveness of intelligent transport systems, Project No. 027763, Deliverable D4.1.1–D6.2, 55
2. Jones WD (2006) Safer driving in the dead of night – infrared vision systems are set to become standard in high-end cars. IEEE Spectrum Green Technology. http://spectrum.ieee.org/green-tech/advanced-cars/safer-driving-in-the-dead-of-night
3. Yendrikhovskij S (1999) Image quality: between science and fiction. In: Proceedings of the IS&T PICS, Savannah, pp 173–178
4. Hertel D, Betts A, Hicks R, ten Brinke M (2008) An adaptive multiple-reset CMOS wide dynamic range imager for automotive vision applications. In: Proceedings of the 2008 IEEE intelligent vehicles symposium, IEEE Press, Piscataway, pp 614–619
5. Hertel D (2010) Extended use of incremental signal-to-noise ratio as reliability criterion for multiple-slope wide-dynamic-range image capture. J Electron Imaging 19(1):011007
6. Marjukka E, Liisa H (2005) MOVE – performance based model for mesopic photometry. Helsinki University of Technology Report No. 35, ISBN 951-22-7566-X, Espoo, Finland
7. Fu C, Li C, Luo MR, Hunt RWG, Pointer MR (2007) Quantifying colour appearance for unrelated colour under photopic and mesopic vision. In: Proceedings of the IS&T 15th color imaging conference, Albuquerque, pp 319–324
8. Hertel D, De Locht C (2010) Utilizing near-infrared and colour information in an automotive wide-dynamic-range night vision system for driver assistance. In: Meyer G, Valldorf J (eds) Advanced microsystems for automotive applications 2010 – smart systems for green cars and safe mobility. Springer, Dordrecht/London/New York, pp 145–154
9. Carlson PJ, Urbanik T (2004) Validation of photometric modeling techniques for retroreflective traffic signs. Transp Res Rec 1862:109–118
10. Keelan BW (2002) Handbook of image quality. Dekker, New York, pp 329–399
11. Hertel D, Chang E (2007) OECF characterization of a non-linear HDR color camera for automotive applications. In: Proceedings of the IS&T 15th color imaging conference, Albuquerque, pp 244–248
12. Dierickx B (2008) Wide dynamic range: what should I care about? IEEE international solid-state circuits conference (ISSCC '08), imager design forum, wide dynamic range imaging, San Francisco
13. Williams D (2003) Debunking specsmanship: progress on ISO/TC42 standards for digital capture imaging performance. In: Proceedings of the IS&T PICS, Rochester, pp 77–81
14. Sangwine SJ, Horne REN (1998) The colour image processing handbook. Springer, New York, p 127
15. International Standard ISO 12232 (1997) Photography – electronic still picture cameras – determination of ISO speed
16. International Standard ISO 14524:1999(E) (1999) Photography – electronic still-picture cameras – methods for measuring opto-electronic conversion functions (OECFs)
17. International Standard ISO 15739:2003(E) (2003) Photography – electronic still-picture imaging – noise measurements
18. Görgens E (1987) Objective evaluation of image quality. J Inf Rec Mater 15:305–321
19. Kerr DA (2008) The ISO definition of the dynamic range of a digital still camera. http://dougkerr.net/Pumpkin/articles/ISO_Dynamic_range.pdf
20. Dainty JC, Shaw R (1974) Image science. Academic Press, London, pp 152–171
21. Chamberlain SG, Lee JPY (1984) A novel wide dynamic range silicon photodetector and linear imaging array. IEEE J Solid-State Circuits SC-19(2):41–48
22. Storm G, Henderson R, Hurwitz JED, Renshaw D, Findlater K, Purcell M (2006) Extended dynamic range from a combined

linear-logarithmic CMOS image sensor. IEEE J Solid-State Circuits 41(9):2095–2106
23. Wang X, Wong W, Hornsey R (2006) A HDR CMOS image sensor with inpixel light-to-frequency conversion. IEEE Trans Electron Devices 53(12):2988–2992
24. Yang D, El Gamal A, Fowler B, Tian H (1999) A 640 × 512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC. IEEE J Solid-State Circuits SC-34(12):1821–1834
25. Kawahito S (2006) A sensitivity and linearity improvement on a 100 dB dynamic range image sensor using a lateral overflow integration capacitor. IEEE J Solid-State Circuits 41(4): 851–858
26. Decker S, McGrath RD, Brehmer K, Sodini CG (1998) A 256 × 256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output. IEEE J Solid-State Circuits SC-33(12):2081–2091
27. Knoll PM (2002) A NIR based system for night vision improvement. SAE Convergence 2002, Detroit
28. Kallhammer J-E (2006) Night vision: requirements and possible roadmap for FIR and NIR systems. SPIE 6198, 61980 F
29. Hörter MH, Stiller C, Koelen C (2009) Hardware and software framework for automotive intelligent lighting. In: Proceedings of the 2009 IEEE intelligent vehicles symposium, Xi'an, China, p 299
30. Hamilton JF, Compton JT (2007) Processing color and panchromatic pixels. US Patent Application 20070024879
31. Honda H, Iida Y, Itoh G, Egawa Y, Seki H (2007) A novel Bayer-like WRGB color filter array for CMOS image sensors. In: Proceedings of the SPIE-IS&T electronic imaging, SPIE, vol 6492, pp 64921 J.1–64921 J.10
32. Kriesel M, Gat N (2008) True-color night vision cameras. In: Proceedings of the SPIE, vol 6540, pp 65400D.1–65400D.10
33. Frame WW (2006) One chip, low light level color camera. US 7,012,643 B2
34. Kidono K, Yoshiki N (2007) Visibility estimation under night-time conditions using a multiband camera. In: Proceedings of the 2007 IEEE intelligent vehicles conference, Istanbul, Turkey, pp 1013–1018
35. Hertel D, Marechal H, Tefera DA, Fan W, Hicks R (2009) A low-cost VIS-NIR true color night vision video system based on a wide dynamic range CMOS imager. In: Proceedings of the 2009 IEEE intelligent vehicles symposium, Xi'an, China, p 273
36. Koyama S et al (2009) A day and night MOS imager spectrally adjusted for a wide range of color temperatures. SPIE 7249, 72490 S
37. Catrysse PB, Wandell BA (2007) Integrated color pixel. U.S. Patent No. 7,248,297 (Filed 2001 – Issued 2007)
38. Choi B (2007) Nano-technology for color: plasmonic nano-optic filter technology for color capturing, expression, and management. In: Proceedings of the IS&T CIC15, Albuquerque
39. International Standard ISO 17321:1998 (1998) Graphic technology and photography – colour target and procedures for the colour characterization of digital still cameras (DSCs). ISO, Geneva, Switzerland
40. Debevec PE, Malik J (1997) Recovering high dynamic range radiance maps from photographs. In: Proceedings of the SIGGRAPH 97 annual conference, Los Angeles, pp 369–378
41. Mitsunga T, Nayar SK (1999) Radiometric self calibration. In: Proceeding of the IEEE computer vision and pattern recognition, Fort Collins
42. Hertel D (2010) Using spectral response measurements to optimise the colour processing of a wide dynamic range colour camera with near-infrared sensitivity. The royal photographic society, symposium on colour imaging 2010, http://www.rps-isg.org/SCI2010_programme.php
43. Scott Tyo J (2007) Displaying advanced optical imagery information for human observers. In: Proceedings of the IS&T CIC15, Albuquerque, pp 174–177
44. Ngo H, Tao L, Zhang M, Livingston A, Asari V (2005) A visibility improvement system for low vision drivers by nonlinear enhancement of fused visible and infrared video. IEEE computer society conference on computer vision and pattern recognition (CVPRW'05) – Workshops, p 25
45. Fredembach C, Süsstrunk S (2008) Colouring the near-infrared. In: Proceedings of the IS&T 16th color imaging conference, Portland, pp 176–182
46. Meylan L, Süsstrunk S (2006) High dynamic range image rendering using a retinex-based adaptive filter. IEEE Trans Image Process 15(9):2820–2830
47. Bertozzi M, Broggi A, Fascioli A (2000) Vision-based intelligent vehicles: state of the art and perspectives. Robot Auton Syst 32:1–16
48. European Commission, Information Society and Media: i2010 intelligent car initiative. http://ec.europa.eu/information_society/activities/intelligentcar/docs/icar/intelligent_car.pdf
49. Carley L (2009) News technologies for safer driving: adaptive lighting, night vision, backup cameras & blind spot detection. Import car, http://www.import-car.com/Article/60749/News_Techn.aspx
50. Kang HR, Anderson PG (1992) Neural network applications to the color scanner and printer calibrations. J Electron Imaging 1:125–134

Books and Reviews

Frieser H (1975) Photographische Informationsaufzeichnung. Focal Press, New York
Hunt RWG (2004) The reproduction of colour. Fountain Press, Tolworth, England
Proceedings of IEEE intelligent vehicles symposium 2007, Istanbul, Turkey, 13–15 June 2007, ISBN 1-4244-1068-1
Proceedings of IEEE intelligent vehicles symposium 2008, Eindhoven, The Netherlands, 4–6 June 2008
Proceedings of IEEE 2009 IEEE intelligent vehicles symposium, Xi'an, China, 3–5 June 2009, ISBN 978-1-4244-3502-9
Reinhard E, Ward G, Pattanaik S, Debevec P (2006) High dynamic range imaging – acquisition, display, and image-based lighting. Morgan Kaufmann Publishers, San Francisco
Sharma G (2003) Digital color imaging. CRC Press, Boca Raton
Wyszecki G, Stiles WS (2000) Color science. Wiley, New York